

From Addison-Wesley Computer Science

KATHLEEN BILLUS
kathleeb@aw.com
kbaw@world.std.com

2.24.94

Brewster,

Tom asked me to send this along
to you.

Kathleen

AAP/PSP

We publishers participate in the creation and dissemination of knowledge. We need to remind ourselves as we go along that our ultimate goal is not necessarily the preservation of publishing as we know it. If we—commercial and not-for-profit publishers, alike—do not clearly understand this, we will lose our role in this process altogether.

This "white paper" is meant to explore some of the issues in the emergence of distributed network publishing. Publishers need this knowledge so they can formulate programs that will insure an ongoing and meaningful role in the process of professional and scholarly communication ...well into the next century. We hope you will find this paper stimulating.

PROMISES & PITFALLS—AN AAP/PSP BRIEFING PAPER ON INTERNET PUBLISHING

Intentionally left Blank

(Inside Front Cover)

DRAFT

Promises & Pitfalls— An AAP/PSP Briefing Paper on INTERNET Publishing

Edited By

Czeslaw Jan Grycz
University Of California
Office of the President

300 Lakeside Drive #8077
Oakland, CA 94612-3550
cjg@dla.ucop.edu

Published by The Electronic Information Committee
of The Association of American Publishers,
Professional and Scholarly Publishers Division,
with contributions from:

Robert Badger	<i>Springer-Verlag</i>
Laura Conley	<i>John Wiley & Sons</i>
James A. Connolly	<i>JEK Associates</i>
Phyllis Hall	<i>IEEE</i>
Stephen L. Haynes	<i>West Publishing Co.</i>
Karen Hunter	<i>Elsevier Science Publishing</i>
John Jenkins	<i>Bureau of National Affairs</i>
Taissa Kusma	<i>American Institute of Physics</i>
Janice Kuta	<i>JEK Associates</i>
Barbara Polansky	<i>American Chemical Society</i>
Blake Ratcliffe	<i>University Microfilms Intl.</i>
Bernard Rous	<i>Association of Computing Machinery</i>
Carol Risher	<i>Assoc. of American Publishers</i>

We also to acknowledge, with gratitude,
contributions from:

Charles W. Bailey, Jr	<i>University Libraries</i> <i>University of Houston</i>
Hans Rocke	<i>University Library</i> <i>Univ. of California at Davis</i>
Win Treese	<i>Cambridge Research Lab</i> <i>Digital Equipment</i>

DRAFT

© 1994 Association of American Publishers /
Professional and Scholarly Publishers Division

TABLE OF CONTENTS

INTRODUCTION

A.	Acknowledgments	1
B.	Distributed Network Publishing	2
C.	Increased Complexity	2
D.	Influence on Policies	3
E.	Value of the Publishers' Imprint	4
F.	Intellectual Property and Responsibilities of Publishers	4

Part I —Overview of the Market

A.	General Trends in the Markets for Professional and Scholarly Publications	7
B.	Trends in the PSP Library Market	9
C.	Trends in Individual Subscriptions	9
D.	Conclusion	11

Part II —Traditional Publishing

Activities	13
Scholarly Communication	
How Publishers Add Value	14
1. ...through the conduct of business	15
2. ...through the administration of their imprint	15
3. ...through presentation enhancement	15
4. ...through increased visibility	16
5. ...through distribution	16
6. ...through leveraging	16
7. ...through legal enforcement of rights and protections	17

AAP/PSP Briefing Paper on INTERNET Publishing

Part III —The Internet —

A Quick Study

A.	The Internet Environment	19
B.	Basic Internet Functions	20
1.	Electronic Mail (E-Mail)	20
2.	Bulletin Boards	21
3.	Telnet	21
4.	FTP (File Transfer Protocol)	21
5.	Gopher	22
6.	World Wide Web (WWW)	22
7.	MOSAIC	22
8.	Wide-Area Information Servers (WAIS)	22
C.	Size and composition of the Internet Market	23
D.	Governance	23
E.	Government Investment in the Information Infrastructure	23
F.	The "Acceptable Use Policy"	24
G.	International Ramifications of the Internet	25

MAJOR ADVANTAGES of distributing information over the INTERNET

A.	Searchability	26
1.	FTP	27
2.	Telnet	27
3.	Gopher	27
4.	WAIS	27
5.	World Wide Web	28
6.	Mosaic	28
B.	Access and Navigation	28
C.	Immediacy	28
D.	Cost	29

PRIMARY DISADVANTAGES of distributing information over the INTERNET

A.	Poor Screen Display Quality	30
B.	Information Quality Control	30
C.	Hidden Real Costs and User Expectations	31
D.	Copyright, Security, and Related Issues	31
E.	Privacy	32
F.	Billing and Transaction Instrumentation	32
G.	Archiving	32
H.	Tenure and Merit	33

Part IV —Migrating to the Electronic

Milieu

File Preparation and Formatting	35
1. Page-Orientation	36
2. Database Mechanisms	37

Part V —Conclusions

Appendix—Case Studies

A.	Journal Fulfillment and Document Delivery Pilots	45
B.	Electronic Journals	50
C.	Publisher/University Pilots	51
D.	Preprints, Table of Contents and Abstracts Databases	53

Glossary

Selected Bibliography

INTRODUCTION

A. ACKNOWLEDGMENTS

This Briefing Paper is a composite of many contributions and results from the work of several individuals. It is likely that the reader will, from time to time, recognize different voices of the various contributors, though some effort has been taken to minimize this distraction by commissioning a single individual to edit the whole.

Having been selected for this task, I am very conscious of the benefit of having had to consider the issues presented by each of the authors of this document. I am grateful for the opportunity to be associated with this text, and hope that whatever modest contribution I may have made in consolidating the several voices into one paper will make it possible for a broader readership to consider the discussion and to advance it through ongoing dialog.

I hasten to take deserved blame for any infelicities of grammar, redundancy, or errors of interpretation that may yet exist in this version of the text. Those faults are entirely mine; the insights, on the other hand, belong to the individual authors.

Most of those who contributed to this final version, are listed the front of this document as members of the Electronic Information Committee of the Professional and Scholarly Publishing Division of the Association of American Publishers. Several of these, notably Laura Conley (John Wiley & Sons), Bernie Rous (Association of Computing Machinery), and Stephen Haynes (West Publishing Co.) con-

tributed extraordinary time and close attention to this paper and deserve special mention.

Worth noting, is the fact that this collaborative writing project was intentionally conducted *on* the Internet. A special ListServ was established, in which all the authors participated. The writing involved a considerable exchange of E-Mail messages over time. The task revealed both the advantages and the limitations of collaborative authoring on the net. It is safe to say the authors will welcome the emergence of any multiple-author document-writing tools that may eventually appear. Such tools are not yet widely available and they were sorely missed. Yet the experience of working together on this document heightened the authors' appreciation for the ability of the network to support a rapid and iterative process such as this.

B. DISTRIBUTED NETWORK PUBLISHING

Distributed electronic networks provide a wealth of opportunity and promise enormous business potential. It is clearly fascinating to the academic and research community, and—by extension—to professional and scholarly publishers. It is not clear, however, whether the network primarily provides an efficient *distribution* mechanism, or whether it exists as a bona fide *publishing* environment.

Much of the confusion surrounding the emergence of "self-publication" on the Internet stems as much from this ambiguity as it does from the apparently widespread misapprehension of publishing's functional role in quality assurance, gatekeeping, and service.

Advocates of electronic networks as the future "communications highways of our society" portray them as already primed for all manner of commercial and publishing exploitation. In truth, there exist substantial barriers which must be overcome before most publishers will feel comfortable about "publishing on the Internet."

The regrettable consequence of this is that there will be a delay in the transfer of professional publishing expertise into the network environment. Publishers—over a period of roughly five centuries—have developed important understandings about the creation, vetting, and processing of publishable ideas. This knowledge could hasten the evolution of the network as a place in which to publish,

rather than merely a place *through* which published materials can travel.

Perhaps the greatest impediment publishers face is the lack of a network infrastructure for the "orderly conduct of business." It may be true that the sociological culture of the net is not entirely conducive to commercial traffic (some would even claim it to be *hostile* to such development), but the absence of controls and protections for normal business transactions is the single greatest impediment to providing the Internet with all manner of scholarly and professional publications in an organized and professional implementation.

It is clear that the integration of business infrastructure is a necessary step for enhanced network robustness. Professional and scholarly publishers are in a position to introduce useful editorial and scholarly communications services, but they are hindered by the lack of guarantees for such ordinary activities as billing, transfer of payments, document security, and policies on privacy.

C. INCREASED COMPLEXITY

That this absence is critical can be understood when one considers the future complexity that publishers will have to manage in a networked environment.

Publishing is a very sophisticated business enterprise. At heart, it is a tapestry of personal relationships, contractual obligations, and responsibilities concerning rights and permissions which have been negotiated against the backdrop of a constitutional endorsement of authors' rights. These rights are presently embodied in the Copyright Act of 1976 (as amended) and in the U.S. implementation of the Berne Convention. Publishing is held together by obligations and by precedents enshrined in a long (at least half a millennium) tradition.

Even in an exclusively print milieu (temporarily setting aside the intensified interactive relationships that characterize network relationships) the obligations between a publisher and contractual partners are frequently difficult to sort. The creation of ever more complex *printed* documents (forerunners of digital compound *electronic* documents) have demonstrated that the complexities (and related capital risks) only have increased over time, as the documents have become more complicated.

For those publishers interested in exploring a networked publishing opportunity, current case law regarding complex print publications can be instructive about the perils, pitfalls, lawful appropriate use, and subtle implications of including illustrations, graphics, and even (seemingly lowly) page numbers in published documents. The current assertion by *Compton's Encyclopedia*, claiming patent rights to a broadly defined suite of hypertext navigational tools, generally comprehended under the term "multimedia" is but one example of the intricacies likely to exist for the publishing business as it migrates to a mainly electronic environment.

Protecting their business investments will be a major preoccupation for publishers who expect to bring electronic information products to market that are

- rights-cleared,
- reliably functioning,
- and
- aesthetically pleasing.

Such "products" will inevitably consist of compound digital source materials; and these will involve ever more complex negotiations over time. (We lack a word in the digital environment as agreeable in its association as is "book" in the print environment.) Compound documents may involve multiple authors, joint ventures, and partnerships with various external entities. They may include sections of programming code, excerpts from databases, and links to remotely located files. Organizing, coordinating, and overseeing these relationships will be a job for an intrepid and experienced publishing professional, else it will risk being managed badly.

And while many professional and scholarly publishers might wish to be excused from such a complex administrative task, it is precisely because of their close affinity to advanced research laboratories, professional societies and institutions of higher education, that they are more likely than most publishers to receive this type of complex document from the authors they seek to attract and serve.

Complexity of this kind suggests an even greater need for well-established legal and business protections and assurances. Such assurances cannot yet be said to exist on the 1994 Internet.

D. INFLUENCE ON POLICIES

Professional and scholarly publishers who venture into the milieu of distributed networks will face another shortcoming. Professional and scholarly publishers have been lulled into a false sense of security, believing themselves to represent a very important but comfortably isolated niche. They think of themselves as masters of their own destinies because of their specialized sphere of influence. Through association with powerful organizations, like the AAP, they have succeeded in lobbying successfully about all manner of policy influencing their business, from postage rate classifications to enforcement of regulations concerning copying of published materials.

But in an electronic environment such autonomy is jeopardized. In the electronic environment there are a lot more voices trying to effect policy and some of them can confuse the issues so its important for AAP to speak with one voice. Publishers are merely one among many, competing for dominance in the dissemination of information to the enormous Internet marketplace. Publishers now find unaccustomed competition from much more powerful businesses. Many of these new competitors are media conglomerates, represent the entertainment industry, or are in partnership with giant multimedia providers: cable operators, telecommunications companies, etc.. These new competitors might well completely overwhelm traditional publishers and their lobbying efforts, and entirely dominate the creation of policy about how information is to be disseminated on the network, who is to have the right to do so, who will be given access, and what economic incentives are to exist for investment in equipment or dissemination channels.

It is more important than ever for professional and scholarly publishers to collaborate with each other and possibly with entirely new partners to ensure a future that includes such widely held (among publishers) virtues as broad and unfettered access to information, freedom and variety of choice, preservation of academic freedoms, and the protection of First Amendment rights.

There may be a beneficial aspect in the crisis faced by PSP professionals, since many will identify closer affinity with librarians, authors, and each other. Inexorably, those who are grounded in the practice of facilitating scholarly communication by creating reliable, vetted, and influential

documents, classifying them, archiving them, and making them retrievable for the benefit of specific disciplines, will find strength in each other more than with the commercial *infotainment* giants.

E. VALUE OF THE PUBLISHERS' IMPRINT

In the face of new competition, it is fair to ask what is of greatest value that professional and scholarly publishers provide their constituencies. While there may be some debate on this subject, one of the most obvious hallmarks of professional and scholarly publishing is the imposition of a reliable *imprint* on manuscripts deemed worthy of publication.

"This is a book published by Wiley (...or West, ...or ACM, ...or (?)...)" means a great deal to the reader or purchaser of a printed publication in the specific fields for which the imprint is a standard of excellence. Imprints—it is safe to predict—will prove to be ever more significant in a strictly electronic environment because they may increasingly serve as the best way for end-users to cull through the undifferentiated chaos of information that courses daily through the network.

Publishers seek assurance that they will be able to impart such imprints meaningfully: without potential future distortion. In an electronic environment making perfect copies is easily accomplished. There exist many means of altering files, but few mechanisms for protecting them. The assurance of the primacy of the imprint is a difficult challenge.

As a consequence, publishers must undertake another educational effort to understand better network architectures, encryption techniques, compression standards, and the technical material that was once the exclusive provenance of the "techies."

F. INTELLECTUAL PROPERTY AND RESPONSIBILITIES OF PUBLISHERS

Finally, some consideration must be given to the publishers' proper care for intellectual properties. In general it is these assets which are licensed to publishers for a certain period of time that permits publishers to generate profit. Their administration of these rights involves responsibilities to a variety of constituents.

...to Authors

Responsibility to the *author*, in whose name materials are released to the public, entails protection of accurate representation of the author's intention and words. It also involves care against misuse, illegal appropriation, or use without proper acknowledgment. It brings the publisher face-to-face with problems associated with copying and plagiarism; and it presumes a capacity to monitor and enforce. Transferring these responsibilities into an electronic environment will clearly add to the publishers' need for mastery over issues concerning encryption, privacy, and data security.

...to the Library Community

Publishers' present responsibility to the *library* marketplace involves establishing and maintaining standards for manufacturing and durability; involves them in proper classification; requires abstracting and indexing professionals; necessitates appropriate recording mechanisms for tracing and retrieval; and mandates appropriate registry and recording of published materials.

In an electronic environment, these needs do not disappear. Rather, they continue to exist as possibly more important requirements. Fulfilling them may be far more difficult since none of the classification *infrastructure* yet exists for electronic document control. Indeed, entirely new tools and mechanisms are yet to evolve for proper document management and retrieval on the Internet.

Those that exist are all in relatively primitive versions at this time, yet they already point to entirely new schema for categorizing, linking, and identifying electronic documents. Librarians are responding to this situation in a variety of ways. Some, for example, systematically produce print-outs of electronic journals to make them *behave* more like their print counterparts; others believe they satisfy their patrons' needs merely by *pointing* to the existence of such documents somewhere on the Internet. Both are endeavoring to establish standardized electronic document [E-Document] controls that will suit the needs of both the library and the networked communities, for such controls are important for files prepared on optical or magnetic media, as well as those intended for electronic transmission over the Internet.

Publishers, accustomed to the well-established routines of print, are legitimately concerned about the lack of standards in this area.

One can predict that there will be a long period of transition, in which electronic and print worlds merge in a variety of different ways, until some successful methods emerge as dominant. For both librarian and publisher the dolorous consequence of this is that—for the foreseeable future—it will be essential to maintain two simultaneous and parallel tracks (one for print, and one for digital documents), with no assurance of added revenues with which to offset the predicted costs.

That the transitory period has begun can be illustrated by a recent posting on the Internet, in which the publisher of the AAAS/OCLC electronic journal, *Online Clinical Trials*, recently shared her great pleasure at being the first exclusively electronic journal to be officially integrated and cited in the major bibliographic resources in the field. (*A Journal's Big Break*. The Chronicle of Higher Education, January 26, 1994)

...to Stockholder

The responsibility to the publishers' *shareholders* involves assuring a reliable return on investment and mandates that careful decisions be made with respect to the technologies and mechanisms that may, in the long-term, prove beneficial to the house and its owners. The quandary presented by this obligation is that while publishers understand the need for transaction software, billing and usage meters, and fulfillment systems, they do not have mastery over how these functions are to be managed within distributed network environments. Nobody does.

...to the Community of Knowledge-workers

The responsibility to the community which publishers serve—especially *professional and scholarly communities*—requires assiduous attention to review procedures, proper editorial controls, maximum distribution to those who can most benefit from the published information, and encourages those who can to use it to leverage increased productivity and research. This is one area in which the network can produce immediate solutions.

The Internet can provide cost-effective solutions to these concerns, especially as network-capable high-speed printers become more widely available. But relying on the network as a means of worldwide distribution again raises issues of copyright protection.

- Electronic disenfranchisement is a possible situation in which those who do not enjoy connectivity with electronic networks may not be able to have access to published literature at all.
- The *national* control of information involves complications arising from the *methods* needed to assert such control, which may make it doubly-difficult for those who cannot pay to receive the information they most require.

It is difficult and expensive for print-based publishers to serve international communities of scholars. Communicating with authors, distributors, and buyers by post takes time, and is frequently unreliable. FAX is relatively expensive. Parcel post is inexpensive but very time-consuming. Transport of physical books to foreign markets is very expensive. Given unfavorable exchange rates, many developing and Third World countries are precluded from participating in the benefit of published information because they can't afford printed materials.

...to Their Heritage

Publishers are also responsible to their *traditions*, one of which includes respect for aesthetics as an important mechanism for enhancing communication. Electronic compound documents today resemble the early experiments with *codex* bound books, at a time when the relatively well-established traditions of scrolls began to be supplanted by folded signatures.

Today's screen displays and linking mechanisms may appear attractive—especially in the minds of the computer specialists who have so skillfully accomplished what appears to most of us as impossible tasks—but they are not as efficient or refined as are the dense and profoundly successful imprints of typography on paper.

Intelligent linkages can span the network linking one "document" (whether text, graphic, audio, or animation) with another. But it is likely that it will take considerable effort (perhaps best contributed by experienced publishing professionals) before visual aesthetics will approach the efficacy of printed materials.

...to Balance

These responsibilities under which publishers labor (or over which they hover) are not always compatible.

The ultimate responsibility of the publisher is to manage the competing priorities of such imperatives so that the reader (end-user) is best served by an editorial process and a delivery mechanism appropriate to the content, at a price consistent with value.

The following sections will explore the issues associated with professional and scholarly publishing in the environment of distributed networks. The observations that follow are intentionally approached from the perspective of the publishing professional; yet they are addressed in a way intended to foster an ongoing exchange of ideas among readers within the publishing profession as well as beyond it.

For despite the barriers suggested above, professional and scholarly publishers have not only expressed a great interest in the new medium available for scholarly and commercial communication, but have already engaged themselves in a series of innovative and sometimes even radical prototype electronic publishing projects (See *Appendix*).

Each pilot project is interesting in its own right, since each challenges certain preconceptions and examines specific possibilities of encompassing electronic opportunities into the present sophisticated and established print experience. Together, they represent an effort to overcome the barriers perceived before the publishers can comfortably take their role in a digital environment. The pilot projects represent a bold exploration of publishing opportunities and constraints. They can be equally as instructive to the Internet community as a whole as they can be to the publishing community which is responsible for them. That is why considerable space is devoted to them in this document.

Czeslaw Jan Grycz
*Scholarship and Technology Study Project
Division of Library Automation
University of California Office of the President*

PART I — OVERVIEW OF THE MARKET

For much of the waning century, the business climate for professional and scholarly publishers has been, if not dazzling, then at least decidedly attractive. All the needed elements for the conduct of profitable and stable businesses engaged in publishing professional and scholarly books, journals, proceedings, and related materials were solidly in place:

- sufficient intellectual resources for new manuscripts and submissions;
- a well-honed and widely appreciated process by which raw manuscripts were vetted, edited, and transformed into formal publications;
- an abundant supply of sophisticated manufacturing technology which could therefore be obtained at competitive prices; and
- a receptive, well-financed, and seemingly inexhaustible market for the resulting publications.

The *market* for professional and scholarly publications is critical for publishers. It is therefore an appropriate subject for the first part of these considerations.

The product of professional and scholarly publishers may roughly be divided into "books" and "journals" (assuming that the variety of conference proceedings, manuals, handbooks, texts, and technical reports can fall into one or the other category of physical format). Professional and scholarly publishers share the market with professional and scholarly societies, most of which have important publishing divisions. This market consists of libraries and individuals who frequently are Society members.

Each of these products demands a different marketing tactic from publishers. Each is characterized by different needs. Changes in priorities, budgets, demographics, or alignments within these markets demand immediate and prompt response from a competitive publisher. So it is worth considering the changes that have recently taken place in the market for professional and scholarly publications.

A. GENERAL TRENDS IN THE MARKETS FOR PROFESSIONAL AND SCHOLARLY PUBLICATIONS

The number of papers and journals published has slightly increased in the past five years. The increase in the amount of overall published research reinforces a need for quality discrimination by traditional quality assurance mechanisms, such as peer review, and endorsement by objective panels of experts... those values that lead to the imposition of the publishers' imprints.

The outpouring of new research enlarges the need for professionals or scholars to keep current with the literature. Indeed, there is some concern that specialization has now fractured some fields, and that the sheer volume of literature inhibits the ability of a scholar or researcher to encompass all the relevant materials in his or her discipline. At the same time, cross-disciplinary use of materials, which is perceived as a fertile and productive intersection among scholars, is also restrained.

This trend has been accompanied by price increases, leading to a growing interest in databases of articles, from which scholars can obtain materials irrespective of the journal in which they are published. This is at the heart of the Document Delivery business, an industry that is thriving in the climate of budget cuts that affect traditional subscription sales.

Unbundling the article from the journal is perceived by many readers to be a beneficial alternative to the printed and bound journal, especially if mechanisms can be designed to implement "fuzzy" matching with which to simulate the kind of browsing that occurs when articles are juxtaposed with one another in a typical bound issue of a journal.

There are risks, however, in pursuing this approach. Article-level "publishing" may reveal that only a small number of certain kinds of articles are ever popular in any given field. There may be a natural tendency to promote the "mega author" which would reduce the fecundity resulting from the availability of a greater variety of thought, even if not all the thoughts are equally impressive.

If pricing models based on a circulation or readership model are adopted, popular articles (which, presumably, would be cheaper, per unit, but would generate greater revenues) would have to support the cost of mounting an entire database of articles that might not carry their fair share of the burden. Under such a "pay-per-view" scheme, costs would be likely to vary greatly. For libraries this is an anathema.

Efforts to level out costs and provide a modicum of predictability might require new site-licensing arrangements, which would tend to reduce article-delivery variability, but would not relieve the cost of mounting a comprehensive database. The site-license would have to provide sufficient revenues to support the publishing operation in the aggregate. Now, however, the basis upon which site licenses

are priced becomes a nasty variable. Does a publisher calculate the potential size of the site, or only its active readership? Do certain sites price themselves out of the market for being too large... too small... too specialized... too generalized?

Publishers are generally cautious about *experimenting* with pricing models, because there is very little elasticity in the market, and it is very difficult to retreat from experimental policies that are seen to fail.

Experiments with putting journals online are equally cautious, for fear that availability of electronic versions will cause a further decline in subscriptions, forcing a premature move towards article-unit delivery schemes before the implications are fully understood.

A number of publishers are (or are considering) creating *customized* journals (not unlike the celebrated PRIMIS model in customized textbook publishing promoted by McGraw-Hill). Under such a scheme the reader selects the papers that will appear in an "issue." The mechanics for this will become easier for publishers when all text and graphics, figures and tables are in electronic format.

The irony is that publishers have long had relationships with typesetting vendors, and have pressured these vendors to become increasingly modernized and computerized. Now, at the point of highest capital investment, perhaps, the availability of relatively low-cost but sophisticated desktop publishing computer software has diminished the loyalty of publisher to typesetter. Similarly, while the publisher has sought to reduce out-of-pocket typesetting costs, the inherent complexity of page layouts for many professional and scholarly journals is such that it requires considerable manual insertion of figures and layout elements. The result is that few publishers are fortunate enough to end up with manageable electronic formats of their publications.

Even those which have converted to an SGML tagging scheme, or something close to that standard, find themselves with retrospective files that would require considerable cost to convert to an electronic medium. This leads to a quandary: Change formats today, causing a logical disruption in the availability of electronic versions of text; or engage in expensive retrospective conversion of files to produce a longer continuum of electronically available materials.

B. TRENDS IN THE PSP LIBRARY MARKET

Libraries face the same conundrum. For electronic text databases to succeed, they must encompass a critical mass of literature that any group of researchers will want to investigate, before that group will be willing to invest the time and energy to understand and integrate electronic text databases into their normal routines of research and bibliographic investigation. This inevitably forces libraries to consider retrospective conversion of publications that are available exclusively in print. But conversions of such files is expensive. Additionally, when publishers learn of such proposals from within the library community, they are alarmed and strive to construct barriers to electronic archives, fearing that electronic availability and dissemination will only aggravate the constricting plight publishers are experiencing.

Libraries which constitute a major market for PSP publications do not, in any case, enjoy surpluses with which to finance expensive retrospective conversions.

Indeed, the budget doldrums into which libraries have been plummeted over the past half decade, are ones from which there seems to exist no quick egress.

Budget cuts have been exacerbated by the general economic recession, which has dramatically affected funding for institutions of higher education. University and college administrators have intuitively perceived a narrowing of distinction between electronic and print-based information on their campuses. While downsizing, some have sought to accelerate the merger of print and electronic information artificially. In some cases computer specialists were put in charge of the academic library, while in others it has been the chief librarian who has been given charge of the computing center.

Whatever mechanisms have been used by administrations to reduce overhead and plant costs on their campuses, library acquisition budgets have invariably been pressured at both ends: by increased costs (especially for serials subscriptions), and by reduced budget allocations from parent institutions.

Neither is the "serials pricing crisis" to which librarians began calling special attention during

the latter half of the decade, an isolated event. And this crisis involves many of the other types of professional and scholarly publications, as well.

The frustrating economic spiral is well-known. As libraries faced increasing serials subscriptions costs, one way in which they sought to compensate was by reallocating some of their monograph acquisition budgets to serials collections. This resulted in fewer monographs purchases, overall. As a result of fewer monograph sales, publishers reduced print runs. But this led to higher unit costs, which, in turn, further reduced the library's purchasing capacity.

These effects, while easy to study in hindsight, took place incrementally over the past decade, compounding the crisis to its present severe proportions. Remediation now is more difficult than it might have been earlier. Meanwhile, both libraries and publishers have received lower "marks" for service by their constituencies, who have pressured for reform, increased cost-effectiveness, and better service from institutions and businesses who are already beleaguered.

The overall result of these cycles and adjustments has been dramatic.

- Funding for monographs has declined to the point where some libraries are actually no longer funding monograph purchases out of standard library operating funds.
- The threat of cancelation of serials subscriptions has become a reality. Even major research institutions—which for so long could be relied upon for consistent buying even in temporary recessionary periods—have now canceled subscriptions even to established and recognized journals.
- The effect on specialized niche publications has been disproportionately heavy, since these publications already had very little margin for profitability.
- Just as the serials/monograph budget cycle results in an economically negative spiral, so, too, has the serials/document delivery cycle resulted in pressure upon both the publishing and the library communities.

C. TRENDS IN INDIVIDUAL SUBSCRIPTIONS

The availability of the network—in the face of these factors—has been viewed as a blessing by in-

dividual researchers and academics, especially within the professional and scientific communities.

Many have explored the network as a mechanism to displace their dependence both on traditional *publishers* and on traditional *libraries*. Without understanding the ramifications (or particularly *caring* about them, in view of what were perceived to be unconscionably long delays in publication, reduced service levels from libraries, and a general belief that both institutions shared Luddite reluctance to embrace new technologies), individual scholars and users of the Internet embarked on all manner of self-publishing endeavors, from launching new electronic journals to mounting electronic repositories of pre-print articles.

It is remarkable that such innovations largely have been initiated, not by the established businesses for whom innovation ought to be an operational priority, but by individuals who had little to lose and much to gain through novel experimentation.

This, too, has an economic basis for understanding. When some wag had it that "Freedom of the Press belongs to those who own one," it could not have been anticipated that an Internet would arise, which—combined with powerful and ubiquitous desktop publishing capabilities—would permit masses of individuals to simulate ownership of printing presses.

The phenomenal growth of the Internet not only exaggerated the rift between the individual scholar and the institutional entities which were established in their support, but also resulted in a phenomenon Lee Jaffe called "drinking at the firehose." So much undifferentiated information flowed through the Internet in so short a period of time, that new mechanisms for classifying, retrieving, archiving, and categorizing information had to be created. Many of these impressive new tools were developed; few from within the tradition and disciplines of librarianship or publishing, much to the detriment of all, but all with attractive features.

Traditionally, individual subscribers enjoyed several advantages when they personally received journals on a regular basis:

- information was current; it was received as soon as issues were published;
- a personal reference shelf could be accumulated; and

- new or ancillary information could be picked up through browsing.

If they relied exclusively on institutional subscriptions they may have had to

- wait for an issue to be available at their library,
- compete with other readers who wanted access to the same issue; and
- wait until a routing process brought them an issue.

In the absence of network capabilities, these advantages were sufficient to maintain a lively sale of individual subscriptions.

In the era of distributed networks, individual subscriptions to print journals, for the most part, have steadily declined as prices for print journals increased. As more information has become available in alternative electronic formats, individuals have adopted new networking retrieval and information management tools. These now include alerting programs by which individuals can be notified of new documents within their sphere of interest. Document delivery services have emerged to serve the need for immediate gratification at costs that can be hidden in other departmental budget lines.

More to the point is that these services have reinforced the utility of the network to help scientists and academic more efficiently perform their work. This inexorable amassing of *service* makes the network increasingly ubiquitous among the natural market segments for scholarly and professional publications. As this trend continues, the attention of our market constituencies will be focused on the Internet. Unless professional and scholarly publishers become vital presences on the network, they will eventually lose the battle for attention.

The nature of individual subscriptions has been subject to the same unpredictable changes that have caused a decline in library subscriptions. As customized products emerge, they will invariably compete for more traditional forms of publication. The real problem is that individuals are more attracted to distributed network solutions for their information needs than they are to traditional forms of publication. Howsoever small that proportion of individuals might presently be, it is a growing number, which will likely continue to grow as institutions and the government continue to invest in the electronic infrastructure.

D. CONCLUSION

The marketplace for professional and scholarly publishers is undergoing profound change in all its manifestations:

- in the instability of the business environment for traditional print-based publishers;
- in increased competition from new and unlikely sources;
- in changing purchasing strategies implemented by various segments of traditional markets;
- in the proportion of sales coming from various market segments;
- in the shifting importance of different products and product lines;
- in pressures from our primary audience for increased service, speed, quality, and technical knowhow; and
- in the diminishing loyalty of traditional audiences.

Arguments can be made about the length of time in which these trends can be expected to manifest really transformatory changes in the fundamental character of professional and scholarly publishing. There are some who will rightly conclude that the changes are so enormous, and that sociological adaptability lags so far behind, that business for print-based publishers will continue to be robust into the 21st century.

But for those who perceive the trends as indicators of an entirely different long-range future for scholarly and professional publishing, an examination might be warranted of traditional publishing activities and values and how these might be modified in an electronic network environment.

PART II — TRADITIONAL PUBLISHING ACTIVITIES

Throughout this century, publishers enjoyed a steady (and steadily increasing) stream of high-quality manuscript submissions. These frequently came from established disciplines. But even emerging disciplines generated their fair share, and did not lack for a cadre of qualified reviewers. The editorial, vetting, and production processes were well-managed, and though each required a professionally capable labor force, there were qualified applicants and considerable competition for any available positions. There existed a sufficiently interested audience for even specialized titles. This was aided by the existence for almost every PSP publication of a core of institutional sales (library standing orders and predictable membership subscriptions), which were a reliable source of financing for even fledgling studies and highly specialized fields. The more speculative individual sales could be pursued with greater confidence when publishers were assured that the entire financial support of a project need not depend exclusively on individual sales.

Advertising revenues were reliable supplementary income for most established and widely-distributed specialized journals, while improvements in print manufacturing technology made short-run printing an economic possibility at the lower end of the distribution scale.

Consequently, professional and scholarly publishers enjoyed a latitude of business opportunities within their principal arena of business: They could exploit the potential of mainstream and already successful publications; or they could satisfy new niche markets and invest in subject areas that might not become economically self-sustaining for some time to come.

All these elements could lead to a characterization of publishing as a "mature" profession. It typically received only incremental modification. This allowed publishers to adapt to changes in the rate or volume of manuscript flow, in advances of typesetting or manufacturing technology, or in the fickleness of audience preference, at a reasoned and comprehensible pace.

Automated typesetting systems, the emergence of desktop publishing, computerized fulfillment systems, and automated distribution warehouses provided an option for some publishers to bring "in-house" functions which had been typically outsourced and paid for by out-of-pocket funds.

At the professional business management level, strategic consolidation of overhead functions and competitive list management made possible the creation of conglomerates which were larger than

typical PSP business units. These would eventually claim highly creditable returns on investment, comparable to those of any other similarly sized enterprises.

Thus, at all levels, and throughout their chosen business milieu, professional and scholarly publishers were in a position throughout the century either to develop new businesses or to profit from existing ones. In such an environment it is easy to understand that while competition among professional and scholarly publishers existed, it was increasingly characterized by specialization and orientation towards specific service areas to which other companies could not easily respond. In other industries, there was a more typical struggle for identical scarce resources or diminishing markets. This was not generally characteristic of the PSP industry.

Some analysts would suggest that this decade, no matter how challenging to individual publishers constituted a comfortable business environment. But they might add "much to the detriment of the profession." Real competitive "edge" is better developed through competition than it is through the less stressful accommodation of dividing the marketplace by specializations. And if, in the closing decade of the century, a number of trends in the marketplace for professional and scholarly products have been disquieting, it may be so because of the loss of competitive "tone" that resulted from this relative comfort.

Today, entirely new mechanisms for distribution and communication have emerged in the form of distributed networks. The government and the public seem to have embraced these new "information highways" as low-cost solutions to information dissemination for all types of published materials serving not only scholarly and professional needs but responding to the needs of the public "from K-Gray", (a reference to the goal of incorporating school materials from the Kindergarten level to public-interest materials that may also serve the aged).

SCHOLARLY COMMUNICATION: HOW PUBLISHERS ADD VALUE

In retrospect, the century can be seen to have been a halcyon period for professional and scholarly publishers, who—together with expert authors and professional librarians—constituted the indispensable components of what we know as "scholarly

communication." Professional and scholarly publishers performed their role with increasing skill and expertise. And their role—coming from an era of arcane mechanics of hot metal type and letterpress printing—was widely recognized for the contributions it made to the efficacy and efficient practice of scholarly communication.

The *process* of scholarly communication was until very recently perceived as a linear progression. An *Author* originated the process, perhaps by expressing an idea in an obscure research paper or proposal. Identified as promising by an astute *Publisher*, a contract might have been proffered and accepted. A subsequently revised manuscript might have been submitted. If so, it would have been shepherded through a process of distillation and refinement which, for some, culminated in a respected and formal publication. Once published, a *Librarian* would ensure that the resulting work would be classified, catalogued, archived, and made retrievable. Frequently, all three of these individuals might be in the employ of a single institution or university. In certain disciplines this is common; in others it is rare.

Regardless of the actual organization charts that can be presented about the structure of the professional and scholarly publishing community, there exists a widespread perception that the *system* of scholarly communication (especially in a digital networked environment) is no longer as well served as it has been in the past (albeit in a print-based environment).

Publishers have long staffed and organized themselves in accord with a linear model of scholarly communication. As the century wore on, each of the partners—author, publisher, and librarian—became increasingly independent from one another, absorbed by somewhat different concerns. Each of these groups tended to face economic and policy problems individualistically. This worked for a time: authors adopted word-processors of various kinds to deal with reduced clerical support occasioned by departmental budget cuts; publishers invested in hardware and software by which to achieve productivity gains; librarians experimented with sharing resources, and developed complex systems of interlibrary loans.

By the current decade, the *system* of scholarly communication had become cumbersome; the economic challenges had become more serious than could easily be resolved by any individual constituency; the linear process approach had become

so extended it lost its elasticity to respond to changes that manifested themselves with accelerating rapidity. Institutions which were home for all three constituencies of scholarly communication reorganized in new and unexpected ways, slashed budgets, and demanded increased productivity and responsiveness. A palpable acrimony arose between authors and their editors over rights, between librarians and publishers over prices, and between universities and their departments over budget allocations.

The overall economic recession also contributed to a reduction in the discretionary spending power of individual scholars. Print runs dropped in reaction to reduced sales. Prices rose accordingly. Libraries canceled subscriptions that made inordinately high price increases.

The downward and complex economic cause-and-effect spiral forms the backdrop against which this White Paper has been written, and requires the reassurance of what it is that publishers really "do" for the community they serve.

Publishers "add value" in several ways to the scholarly communication process.

1....through the conduct of business

First among these, in the minds of many of the workers within publishing houses, is to manage the hassle, detail, and responsibility involved with maintaining records, handling payments, and managing business records associated with royalty payments, permissions, and contractual obligation.

This may come as a bit of a surprise to non-publishers, but will be readily understandable by any publisher. The process of handling and managing records for transactions, payments to vendors, and service providers, royalties to author(s), and fees to various agencies is not insignificant. These transactional costs might be borne by some more automated software agent. To date, however, there has been more discussion about such agents than there have been examples of working prototypes, especially those that could be generalized and have the sufficient scalability to work in distributed network environments.

2....through the administration of their imprint

As has already been stated, the imposition of an *Imprint* by a publisher is a form of endorsement which is all the more significant for professional and scholarly publishers, because of the objective

forms of blind and expert evaluation that is variously employed to assure the publisher that a submitted work is worthy of attention and suitable for publication.

Many of the functions performed in this capacity seem modest and practical. Together they add up to an important value contributed by publishers, because their imprint is one of the few ways of differentiating among varying levels of importance of published materials. These functions include

- provision of formats/models which authors can follow; standardization provides quality control;
- filtering of material internally (publishers' editorial evaluation) and/or externally (by peer reviewers);
- provision of screens against low quality, incorrect, or libelous publications;
- refining of content and context through such review, as well as through the work of professional editorial staff; and
- certification of publishable materials which have successfully passed through such a process.

Any publisher who submits manuscripts to review will agree that incoming manuscripts frequently bear little resemblance to publishable books and journals. Often an important transformation takes place merely because of the *process* through which it is passed.

3....through presentation enhancement

Similarly, intellectually compelling ideas sometimes can be enhanced through appropriate presentation in articles, books, or other publications. Our constituents are not only literate, they are—as part of our cultural inheritance—sensitive to subtle aesthetic signals in the presentation of published materials. Those signals can influence the inclination of a reader to absorb or understand especially complicated materials. These qualities might be added through

- quality control in copyediting, typesetting, artwork preparation, layout, formatting, printing;
- packaging in widely accepted, "reader friendly" formats; or
- improved clarity of ideas through various forms of visualization of complexity, as in tab-

ular arrangement, position, or alternate representation.

4....through increased visibility

It has been asserted that the two indispensable activities which define a publisher are *selection* and *dissemination*. If the former two activities have to do with proper selection of manuscripts, and the appropriate response to those that are meritorious, the next has to do with dissemination. Marketing and advertising departments often seem superfluous to network enthusiasts. Yet, the increased visibility given to publishable manuscripts by these publishing departments has much to do with a publication reaching its intended audience.

Without these critical activities, even the best manuscript can languish without sufficient readership:

- maintenance of current customer profiles and interests/Market Research/Circulation Management;
- extensive marketing and promotion (direct mail, database marketing, catalogues, advertising, and exhibitions);
- maintenance of Current Awareness services (databases, print products, ISI, etc.);
- support of secondary publishing venues such as abstracting and indexing;
- provision of customer service departments

What distinguishes these forms of added value is their *action*. Marketing and Advertising departments are not passive, but seek to expand the readership beyond the natural one that exists for any specialized manuscript. These efforts have a profound effect on reader awareness, the salability of books and journals, and their beauty as vehicles of information.

5....through distribution

Books and journals need to reach their intended audience. A variety of specialized services and tasks support this capability, and a variety of options have also emerged as publishers have attempted to be more responsive to their customers' needs, within the limits imposed by the physicality of printed materials

- rapid publication and use of cost effective distribution networks, (e.g. Publishing Expediting Service);
- investment in worldwide sales infrastructure (sales offices, sales representatives, book-sellers, book shops, subscription agents, and maintained contact with customers);
- provision of reliable and known channels through which to obtain publications;
- an attractive series of discounts to promote the availability of printed materials in various bookstores and appropriate outlets;
- creation of book clubs and other enhancements through which individuals can obtain reduced-cost materials in their sphere of interest; and
- archival availability of past products (accomplished through cooperative library deposit programs) and sales to that channel.

6....through leveraging

Publishers become recognized in their various fields through their successes in attracting the best authors, providing a critical mass of titles on a list, and becoming *the* source of information about their specialization. This doesn't happen automatically, but involves a host of interrelated activities not earlier identified

- planning of a portfolio of publications in response to market trends and needs monitored through market research/market assessment;
- initiation of publications that help define interest groups and subject areas of research;
- encouragement of a tenure and merit approval process that gives sufficient merit to certified publications;
- extended and ancillary activities that facilitate communication among scientists and authors, e.g., organization of conferences, symposia, seminars, one-to-one contacts, and small groups (journal editorial boards, author groups); and
- funding of research grants, student awards, and competitions for excellence.

7...through legal enforcement of rights and protections

Earlier, it was noted that the publishing activity is supported by a complex code of rights. No property can survive if it is not defended against misuse. In their capacity as agents for authors, publishers have taken seriously their obligation to defend copyright, enforce intellectual property rights, and prosecute plagiarism or unauthorized use of published materials. They promise

- protection of Copyright on behalf of author and publisher;
- initiation and negotiation of license and rights options (foreign language rights, rights of anthologies, electronic rights, etc.);
- centralization and control of copying through the Copyright Clearance Center;
- payment of revenue share/royalty to authors; and
- formulation of quality and policy control boards to guide industry practices.

These constitute a suite of services and values of which professional and scholarly publishers should be proud. There is no doubt that the vitality of scholarly communication is to a large degree the product of the professionalism of professional and scholarly publishers in fulfilling these commitments.

DRAFT

PART III — THE INTERNET: A QUICK STUDY

Professional and scholarly publishers have faced huge market pressures in a condensed period of a few short intense years. Rapid technological changes have forced publishers to increase capital and operational expenses for equipment, software, and training. These significant expenses have been justified by the promise of expanding circulation and sales. But analysts generally predicted a flat, if not declining, market for professional and scholarly publications.

The audience for PSP publications is largely university and research communities based which have had access to the rapidly expanding distributed network and have more experience with its potential than any other market segment. While enthusiasm is boundless among academics and researchers for the advantages of collaborative workgroup functionality, distant interaction with colleagues, shared resources, and remote job control, it must be admitted that the practical aspects of running a business are not their primary concerns.

Nevertheless, they have incontrovertible evidence of the utility of distributed networks to facilitate the advance of research and scholarship. We

serve those constituencies and need to assure our patrons that the services and values they receive in print will also be provided in the electronic milieu.

A. THE INTERNET ENVIRONMENT

The Internet, as all-pervasive and seemingly ubiquitous as it appears to be today, is actually a technology of relatively recent provenance. Begun by the Advanced Research Projects Agency (ARPA) and other federal agencies in the late 1960's, the network's original purpose was to permit persons working on federal research projects anywhere in the U.S. to tie into computers at remote sites, use those computing resources, obtain files stored at those computers, and collaborate electronically with scientists and other personnel across the country. Today the Internet is the supranational information highway connecting networks of federal, regional, campus/academic, private and foreign users.

At the simplest level, the Internet is simply a "network of networks" a vast web of wires, fibers, and microwave circuits connecting thousands of physical networks, linking millions of host computers and tens of millions of users worldwide. The Internet has now blossomed in the national consciousness both as an example of successful high technology and as the current home of "hackers" and "computer nerds." Functions of the Internet—which will be discussed below—have become so appealing and, in fact, indispensable to its millions of users that usage of the network has been growing at the compounded rate of 10-15% per month (!) for

the last few years. This growth rate shows no sign of abating. It is growing with equal rapidity in Asia and Europe, and even in Central and Eastern Europe, and—at a slightly lower rate of speed—in the various republics of the former Soviet Union. It is, indeed, a global information infrastructure, though national consciousness makes it appear to be primarily a national asset.

The Internet promises soon to be the electronic foundation of the National Information Infrastructure and the High-Performance Computing and Communications superhighway about which the present administration is so enthusiastic.

The principal use of the Internet is to transfer files from one computer to another. Thus, an author at the University of Michigan may send to her publisher in New York City a copy of her manuscript (e.g. in WordPerfect format) via the Internet. In fact, the entire process of initial submission, review with comments, submission of revised drafts, editorial markup, and submission of a final draft can easily take place on the Internet, and some publishers are doing so, though most are experimenting in a somewhat haphazard manner.

The second greatest use of the Internet is for electronic mail. Here, the author and publisher confer in the electronic equivalent of an exchange of letters. As is discussed below, E-Mail has myriad other uses as well.

Evolution of the Internet has included steady increases in transmission speeds, increases now measured in orders of magnitude. At one time, 56,000 bits per second (roughly 5,000 characters per second) was the topmost speed on the Internet. In relatively recent times, speeds of 1.5, 3.0, and 4.5 million bits per second have become commonplace. Some installations now support (in excess of) 40 million bits per second, and the goal is for the Internet routinely to support (in excess of) 1 billion bits per second.

This gigabit-per-second milestone is a goal of legislation that Congress passed two years ago. At those speeds, audio, graphics, and even full-motion video transmissions through the network will become feasible. At a gigabit-per-second transmission rate, the entire *Encyclopedia Britannica* may be transmitted from New York to Sydney, Australia, in significantly less than a minute. We are on the verge of a revolution in access to and use of information. Such rates of speed make possible all manner of delivery mechanisms (e.g., On Demand book-

stores are actually feasible) that were only recently the stuff of Buck Rogers.

To participate in the spread of Internet-based information resources, publishers will have to understand its potential for good or ill.

B. BASIC INTERNET FUNCTIONS

1. Electronic Mail (E-Mail)

a. Overview

E-Mail, as the name suggests, allows Internet users to send and receive mail to and from others on the Internet. Data files and executable files may be transmitted over the Internet as forms or attachments to E-Mail messages.

As with the E-Mail systems typically in use at any publisher's office, Internet users can send messages or files to anyone, anywhere on the network. Messages are stored until they are opened. (Store and Forward), which relieves the requirement of sender and receiver to be "present" to a communication simultaneously. Access to Internet E-Mail from many internal or commercial E-Mail systems is also available via various gateways (i.e., local or commercial networks which have provided links with the Internet backbone).

b. How to send E-Mail

To send a message, a user must know the address of the computer account of the desired recipient, and the name of the computer ("host") on which the account exists. The following is the standard form for E-Mail addresses:

- [user name]@[host name].[domain name].[account type]

On many Internet hosts, to send a message, a user enters the word "mail" followed by the desired address. From that point, the user is prompted to enter the subject, the text of the message, and, sometimes, the addresses of copy recipients. Received messages may be forwarded to others on the Internet.

Individual commands by which to create, send, receive, file, attach, and manage E-Mail are increasingly hidden from the user by any of a variety of elegant *interface* programs, which facilitate all these operations in a "user friendly" way. As these have come along, the increase in the use of E-Mail

has been dramatic. No longer is there a learning barrier to the utility of the Internet.

c. ListSers

ListSers are software programs that enable and maintain E-Mail discussion groups. Each of these discussion groups is dedicated to some specific topic (often indicated by the name of the list). ListServ software keeps track of subscribing account-holders and reflects to each subscriber messages sent to a central source. They act as distribution centers for mail messages. Many ListServ groups are gatewayed to USENET newsgroups (see below). Typically, to subscribe to a ListServ list, one sends an E-Mail message to the ListServ located on the same computer host as is the discussion group one wants. The subscription request is sent as a single line, conforming to the standard

<SUB Wonderful-List YourFirstName YourLastName>

2. Bulletin Boards

a. Bulletin Board System (BBS)

BBS is a tool for providing bulletin board services. Any given BBS will offer a combination of E-Mail services and collections of data and/or documents that are available for downloading. A number of BBS services are connected directly to the Internet. Many bulletin board systems are operated by government, academic and research entities. While often offering E-Mail, bulletin boards are different in that messages are not collected by the user's local computer but rather are stored on the BBS until accessed by the user. BBSs will often offer "fora" in which users may post messages directed at others having the same interest as that addressed by the forum—e.g., fora exist for stamp collecting enthusiasts, the use of Microsoft Windows, or for discussing the current hot political issues.

Other users may selectively access message "threads" (groupings of messages and responses made to those messages along topical lines) based upon whether the subject of the thread sounds interesting. These fora permit a "gentler" form of information dissemination than do ListSers, because it is up to the user to collect mail, rather than receiving it simply by virtue of having subscribed to a list. Frequently, list discussions can become very energetic, and the volume of incoming mail can increase dramatically if one subscribes to one or more lists.

Non-Internet BBSs often involve a fee for access.

b. USENET (Users Network)

USENET is an Internet-restricted electronic bulletin board network comprised of thousands of topically divided "newsgroups." Each newsgroup will appeal to a limited subset of Internet users. Some newsgroups may have literally thousands of subscribers while others may exist with only a dozen or so subscribers.

The economics of electronic information distribution is nowhere more apparent than in the economics of such groups. For it costs no more or no less to mount and maintain a thousand-member group than it does for a seventy-member group. It is this independence of unit cost variability that is one of the more appealing aspects of network distribution.

And access to USENET newsgroups is free.

3. Telnet

Telnet is the primary Internet protocol used to log in to a remote computer and use its applications. Telnet allows users to be on one computer system and do work on another, which may be across the country or across the world.

This has been an enormous benefit to various forms of scientific research which depend on extremely powerful (therefore expensive) supercomputers. By networking such valuable resources, the entire scientific community benefits from eliminating redundancy by maximizing the computing cycles on a given computer.

4. FTP (File Transfer Protocol)

a. Overview

FTP permits the retrieval of files from remote computers. FTP is also the name of the program that implements the protocol. While the use of FTP requires a user id, the "anonymous FTP" service allows users without passwords to enjoy limited access to some computers. Various FTP sites have been established across the network, consisting of computers dedicated to maintaining file archives for anonymous FTP.

Individuals can retrieve text files, application programs, binary code, software updates, and various utilities and computing aids in this manner. Many *Internauts* find FTP sites attractive augmentation to publishing, presuming that if publishers would only mount their journals and articles on FTP

servers, the entire community of scholars would thus benefit.

Of course, an FTP site commonly takes no responsibility for the accuracy or operability of the software mounted on it. Indeed, many of the sites are operated by the very user community which uploads and downloads files to it. The exchange mechanism is based on altruism and good-will. As a business model, it is an attractive ideal, but it lacks a certain amount of practical implementation for the recovery of costs for those files and applications which have involved significant capital or time investments.

b. Archie

Archie locates files at anonymous FTP sites by filename search. Archie was created to track the contents of anonymous FTP archive sites.

It functions by polling predetermined FTP sites and gathering the filenames of the files presently extant at the sites. It then sorts and captures the filenames in a way that permits keyword and title searching. The "hits" reveal the location of the original file, permitting the user to connect with the appropriate host to retrieve a desired file.

5. Gopher

a. Overview

Gopher is a point and click menu-driven access tool developed at the University of Minnesota. Gopher permits Internet users to access data without having to know beforehand the location or details of what they are seeking. Prior to the creation of Gophers Internet users had to know exactly where and by what name the desired resources were hidden. Effective use of the Internet required considerable experience and a great deal of communication among users about where resources were located. The Gopher provides a simple, menu-driven, tree-structured interface to the data/files available on the various Internet servers.

Gophers are also associated in families. Once you connect to one Gopher, you enjoy seamless access to every other Gopher, making navigation—at least among this class of resource—remarkably easy and intuitive.

b. Veronica

Veronica is a Gopher service dedicated to locating resources existing within "Gopher Space." The

Veronica service maintains an index of titles of gopher items and provides for keyword searching of those titles. The result of a Veronica search is a gopher-like menu of files or directories.

6. World Wide Web (WWW)

WWW is a network navigator similar to Gopher. It is a menu-driven Internet navigation tool. Unlike Gopher, WWW relies on hypertext links between related sources. As a consequence, a scholarly paper on one Internet-connected computer can incorporate citations to other papers or resources found elsewhere on the network; a user reading the primary source can then link via the hypertext references to the other resources and WWW will automatically connect them to the cited resource. WWW allows users to travel the "strands" of a web locating information of interest. WWW promises immense potential and/or threat to professional and scholarly publishers.

7. MOSAIC

In anticipation of an increasing desire for compound documents from the network, a suite of Internet network tools has been gathered under the rubric "Mosaic." These include a variety of visualization tools that work with UNIX, X-Windows, Macintosh, and other platforms to permit the display of graphic images. These can be static or animated, black-and-white or color. The user is relieved of the burden of identifying the appropriate tool with which to display a file, since the Mosaic suite is capable of selecting the appropriate software for a variety of file formats.

8. Wide-Area Information Servers (WAIS)

WAIS lets users of various operating systems access databases across multiple Internet servers. WAIS is a tool for searching word-indexed databases, and its search engines accept natural language queries. Queries are treated as a string of individual terms or phrases. The system reads the user's query, searches the database(s) for documents containing the requested terms and phrases, and ranks the results using a term-weighting algorithm to generate a relevance list. Documents are ranked as a function of how likely they are to conform to a user's expectations.

The document that best matches a search is assigned a relevance score of 1,000 (the remainder of the retrieved documents receive lower scores).

WAIS allows users to refine their searches using relevance feedback. Relevant documents may be selected and sent back to the server to further refine a search. ("Get me more like this.") Hundreds of WAIS-accessible databases now run on the Internet.

C. SIZE AND COMPOSITION OF THE INTERNET MARKET

During its early years, the Internet user community numbered only a handful of agencies and research centers. In 1986, the National Science Foundation took over primary responsibility for managing the network. At first, the NSF saw the Internet as a way to provide general access to rare and expensive scientific instruments like supercomputers. Soon, however, the network evolved a life of its own as users logged-on to begin discussion groups, send e-mail, and conduct research.

Under the NSF's care, the Internet has grown at a phenomenal rate. It now includes 15,000 subnets embracing almost two million computers and probably well over three million users (some estimates place it as high as 15 million users) in 175 countries.

While the user community was once dominated by government, usage of the network has broadened in recent years. A recent study revealed that government and defense traffic now constitute less than a fifth of the information passing through the network:

- 48% research (including commercial)
- 29% commercial
- 10% defense
- 7% government
- 6% educational

Approximately 6.75 terabytes of data were transported on NSFNET (the backbone of the Internet) in July 1993. This implies nearly 100 terabytes of data in the total NSFNET system that one month—or the equivalent of the entire Library of Congress being sent every month.

D. GOVERNANCE

Internet management is highly decentralized. There is no single, authoritative governing body. Instead, several agencies work cooperatively to manage the system:

- The Internet Society (ISOC) is a voluntary organization which promotes global information exchange through the Internet.
- The Internet Architecture Board (IAB), appointed by the ISOC, allocates network addresses, focuses problems, encourages standards and maintain records.
- The Internet Engineering Task Force (IETF) works with the IAB to consider operational and technical problems. Often it will appoint working groups to investigate specific issues.

Even with these agencies to support and guide Internet growth, local, regional, national, and international traffic pools, pipes, and policies come into play in a sometimes anarchistic manner. Each user community may have its own set of rules. Etiquette, defined and monitored by the user community—as much as anything—regulates how users interact on the Internet.

Paul Evan Peters, Executive Director of the Coalition for Networked Information (CNI) provides a useful analogy when he describes this seeming anarchy of the Internet: Drawing from evolutionary textbooks, he likens this stage of development of the Internet to the "Paleolithic period." "It is an environment," Peters says, "in which crude tools are being used to fashion crude, but functional artifacts; in which the dominant personalities are hunter-gatherers and story-tellers; in which institutions and organizations, including libraries and information centers and providers of all types, are hard at work securing the gains of these pioneers by constructing fixed settlements that are attractive to settlers who are much more interested in husbanding domesticated flora (i.e., databases) and fauna (i.e., algorithms) than they are in exploring what's over the next horizon."

E. GOVERNMENT INVESTMENT IN THE INFORMATION INFRASTRUCTURE

It is difficult to track government spending on the Internet. Direct spending seems surprisingly small: \$17M was allocated to the NSFnet for 1993. However, the U.S. government has supported the Internet primarily through "hidden" investment: the Department of Defense, NASA, major research universities, and other related national expenditures were estimated at over \$300M.

Outside the U.S. it is much the same story: The UK has spent millions of pounds to develop the JANET; Japan is aggressively developing NACSIS; and the French government has poured money into their information infrastructure. All developed nations have followed the example, and all these networks will join each other via the Internet.

Similarly, developing countries perceive the network as the most economical means of accessing needed engineering, environmental, and even political information. Those countries, too, are investing heavily in the one resource that is truly trans-national and takes the form of pure democracy, in which every user is every other user's equal, no matter what her skin color, economic status, or social achievement.

Government investment has had an enormous impact on Internet growth, supporting and nourishing it during a time when commercial use (and thus investment) was prohibited. Two years ago, the NSF lifted this restriction on commercial traffic and investment from private agencies has increased dramatically.

Today, the U.S. government investments are rapidly moving from the "supply" to the "demand" side of the total funding equation. Most major users pay for a piece of the Internet through reciprocal accounting agreements. Federal agencies spend tens of millions of dollars each year to build and maintain network infrastructure; colleges and universities contribute to campus and regional networks often via state and federal funding. Commercial providers maintain their own segments of the network hardware and services and sell access to others. Large companies may pay tens of thousands of dollars per year for Internet access; smaller companies and institution pay lesser amounts.

F. THE "ACCEPTABLE USE POLICY"

During the first two decades of its existence, the Internet prospered and grew with the help of substantial federal government subsidies. Especially as regarded the high-speed "backbone" of the Internet—that portion intended to connect the primary supercomputer research centers—federal monies played a significant role, first in the research that brought about the backbone, and then in its implementation. For several years now the Internet backbone, under the supervision of the National Science Foundation, has been subject to an NSF-promulgated "Acceptable Use Policy," or AUP.

The AUP essentially forbade use of the Internet (or, most particularly, any federally subsidized portion of the Internet) for commercial purposes. It stated this *proactively*, by mandating that the federally subsidized portions were to be used only for "education and research."

In 1992, however, the company Advanced Network & Services, jointly created by Merit Corp., IBM, and MCI to serve as the principal contract agent for Internet services, set up a subsidiary, ANS CO+RE to promote the commercial use of the Internet. ANS CO+RE first had to circumvent the AUP.

ANS CO+RE devised an elegant way of doing so by structuring the "routers" on the Internet (those computers at the network nodes that route Internet traffic) so that traffic originating at or destined for commercial addresses (i.e., those with ".COM" as the last suffix of the address) was sent by channels *other than* across the NSF-subsidized backbone.

As a consequence, commercial entities on the Internet may be assured that they can offer virtually any service on the Internet and not run afoul of the AUP.

In addition, the NSF itself was directed by a 1992 amendment to the National Science Foundation Act to encourage use by the education and research communities of non-AUP-bound networks (i.e., non-AUP-bound parts of the Internet) when those networks "may be used substantially for purposes in addition to research and education in the sciences and engineering, if the additional uses will tend to increase the overall capabilities of the networks to support such research and education activities." 42 USCA 1862(g). Thus, the NSF is itself no longer strictly adhering to AUP principles.

For those interested in moving traditional publishing onto the Internet, however, the special provisions for .COM addresses (above) implies a corollary for those who may have accounts with suffixes like ".EDU" (educational institutions) or ".ORG" (societies and non-profits). Materials sent to them or the work of their editors, to the extent that work/product is addressed to or originates at such non-.COM addresses, may now be violating the AUP.

Partly to avoid such paradoxes, and partly to fulfill the mandate proposed by the Clinton Administration, NSF has recently announced the creation of a series of POA (Points of Access), which will permit commercial providers to access the

network without violating the backbone AUP restrictions. This is the first stage in commercializing the Internet. While this decision comes as a disappointment to the current users of Internet (many of whom would prefer to operate in an idealized, somewhat anarchic, and very altruistic environment), it signals a new evolution of the network, in which commercial traffic and tools to support such activities are likely to begin emerging.

It is unclear whether network distribution will remain free. There are certainly real costs in setting up and maintaining publication servers. Right now, these costs are often hidden because an institution may support such activities and users never directly realize the cost. Also, the infrastructure satisfies many universal needs for communication and information access, besides publication in a stricter sense. So even if the user paid the real costs, they would be universally distributed and would be quite low.

Perhaps tuition will rise to cover the large university investments in the technology infrastructure and personnel to operate and maintain it. But individual faculty who may volunteer their time to manage a particular information resource have their own labor presumably absorbed in their salaries. The net result is that end-users see the information as free (and often cannot understand why publishers want to charge). And local printing costs of equipment and paper are also usually hidden.

Now it is probably true that if publishers can stop printing completely and simply distribute the electronic page files, they will probably ultimately lower their own costs. Additional benefits might be provision of more articles once the page budget related to press, binding, and distribution costs disappears.

G. INTERNATIONAL RAMIFICATIONS OF THE INTERNET

It is said that we have entered the "information age" and that the future economic health of the United States is closely associated with its intellectual properties which will be in much demand over the next decade. Many of the world's scientific discoveries, Nobel prize winners, and significant medical achievements are fostered in this country.

The international dimensions of the network have already been suggested. Electronic delivery

and sale of information could have ramifications for the economic well-being of publishers, as well as for the communities beyond our borders who will want to retrieve and access information generated by U.S. institutions and publishing businesses.

The Internet permits the consideration of a global market, global distributed manufacturing, and global rights exchanges.

As a professional and scholarly publishers do we license translations or co-editions? If a library or other customer could order the same material either locally or from the primary publisher; would this impact on foreign rights?

Access to the Internet is available in 134 countries. Internet communication is preferable to FAX and telephone at universities in Estonia and Ukraine, two countries where the telecommunications infrastructure is grossly inefficient. Although telephone lines frequently break-up, and FAX transmissions, as a result, are more often incomplete, the Internet and its satellite hook-up permit reliable communication by E-Mail with people around the world.

As with computers, Internet usage only grows as people learn to use and depend on its efficiencies. Today, libraries use their international E-Mail to request interlibrary loan copies of printed materials from other libraries, requests that may be filled by mail. Soon it may be expected that such requests will be filled by FAX transmission or over the network itself. When a library patron needs material that is not available in a library's holdings, securing that material quickly becomes a priority.

Timeliness is one of the great advantages of the Internet. Minutes after the devastating earthquake in Los Angeles in January '94, the Internet stayed up when other communications systems failed. Disaster Relief agencies used the Internet to provide information about closed roadways, damaged structures, and relief sites. People traveling in foreign countries received confirmation from their families that they were well and unscathed. This kind of immediacy may not be relevant to a publications process that depends on careful deliberation, but it is one more signal that the Internet is becoming a communications channel of wide-ranging utility.

Libraries are not insensitive to the advantages of such a communications link. Whether an interlibrary loan request is filled by a library making a copy from its print holdings, or by a document de-

livery service, or by the original publisher, becomes merely a function of speed, cost, availability, and ease of locating the materials.

Instead of waiting for a request from a patron, some publishers will be offering their materials over the network immediately.

The network, from this perspective, becomes a better form of distribution system and one with the most comprehensive reach imaginable. By virtue of search engines, anyone in the world with access to the Internet can find the pertinent materials.

Today, the Research Libraries Group's *Ariel*TM software is being tested in England on the SuperJANET (Joint Academic Network) for interlibrary loan and document delivery of scanned page images. One anticipated result of the experiment is an increased demand for network delivery of published materials.

Both the British Library and the French National Library (Bibliothèque Nationale) are using the Internet to access the Library of Congress electronic catalog. Document delivery is the next logical step.

Because the network has no logical geographic boundaries, publishers might wish to consider network publishing as a global distribution system with all the attributes that such would entail in a print environment, including the need to insure that world rights are obtained for copyrighted materials included in a network offering.

Of course collecting and distributing compensation for such international delivery raises an entirely new set of questions. Recent legislation in Taiwan resulted in a requirement that only copyrighted materials licensed for Taiwan be permitted to enter the country. This closed-borders policy caused a new inquiry into the comparative pricing of various serial publications available in Taiwanese editions. An international distribution system will subject the publisher to the same type of inquiry.

MAJOR ADVANTAGES of distributing information over the INTERNET

Because both text and graphics may easily be transmitted via the Internet, a publisher might require that all articles be submitted to it via the Internet. This would have the additional advantage of forcing a certain standard (ASCII), although—as stated previously—it is entirely possible to transmit files conforming to any format or binary file representation.

Materials or software the publisher makes available to its authors might be stored on the publisher's host and downloaded by authors using FTP. Any new author might have access not only to stock materials provided by the publisher, but to an archive of past articles, including those closely related to the author's subject, possibly searchable using WAIS(es). When utilized properly, the Internet should be a major boon to the publisher's efficiency, decentralized operation, and timeliness.

What is needed is a creative embrace of the various existing tools and an orderly structure for ease-of-use.

The possible publishing uses of the Internet are legion. As one very real example, this White Paper was prepared, collaboratively, by nearly a dozen people each working at his or her separate company. Various pieces were exchanged using the Internet, comments and amendments were submitted across the network, and the final document was the joint product of people working literally on both coasts. Whereas the few meetings that were physically held to discuss the paper took place at AAP offices in New York, the computer through which the Internet collaboration took place resided at the University of California at Berkeley.

But the distributed network environment provides a myriad of advantages for file access and retrieval, any one of which will be of interest to an entrepreneurially oriented professional or scholarly publisher:

A. SEARCHABILITY

The Internet offers searchability of very large bodies of information (efficient intellectual processing); hypertext; manipulation; data integra-

tion; Selective Dissemination of Information (SDI); more-than-text (e.g., audio, video, datastreams); etc.

Because of its speed and the ubiquitous nature of its underlying TCP/IP protocol, the Internet may be considered a single, large, distributed computer system. Since a user in Minneapolis, Minnesota with the appropriate authority and passwords can sign onto a computer in Sydney, Australia and for a period of time use that computer's resources as if it were just down the hall, the opportunities for distributed work and distributed resources are immense. In a similar manner, information may be widely distributed over the Internet, yet tied together as a single intellectual resource.

Ease of access to this information may be viewed as stratified according to the nature of the tool used: FTP, Telnet, Gopher, WAIS, World Wide Web, or (most recently) Mosaic. Some of these tools have been described at some length above, so this discussion will concentrate on their utility for managing information in ways publishers might find interesting.

1. FTP

File Transfer Protocol simply allows one to sign onto a remote computer and acquire (download) files stored on that machine. The files may be in ASCII or binary form (i.e., they may be images as well as text, or for that matter may be executable programs). In this sense such information is distributed, but you need to know precisely where to go and precisely what you are seeking.

- Publishers could easily establish files of instructions, standard contracts, editorial style guides, and application programs to permit tagging and conforming to publishers' predilections.
- They could easily set up archives of articles both prior to publication and after.
- They could mount electronic versions of files which have appeared in print for over a year without greatly threatening the current subscription level.

2. Telnet

Telnet capability normally does not deal in information management. However, since using Telnet capability means not only obtaining access to a remote computer but using that computer's facilities, those facilities could include information search

and retrieval capabilities. Thus, if the remote computer had a searchable database, the Telnet-ing individual could gain access to that database and its search engine.

Moreover, if a normal user of the remote computer were so foolish as to leave in an accessible file his or her password to DIALOG or WESTLAW, the Telnet-ing individual could use telecommunications resources on the target computer to access one of these online services and perform searches that would cost the target facility.

Unlikely as these possibilities may at first appear, such abuses of the Telnet-ing capability have, in fact, occurred and are one reason for concern about Internet and Internet facilities security.

Still, the obverse is true, and a clever publisher could provide search and retrieval software to its own list of titles to permit browsing, identifying specializations, and even permitting book sales.

3. Gopher

Gophers may be viewed both as finding tools and road maps to information resources on the Internet. Veronica, the search tool associated with Gopher, may be used to locate places where information is likely to be found without having to traverse the entire Gopher tunnel structure. ("Tree structure" [except that gophers don't climb trees]).

One can create and maintain a wealth of information on the host at an end-branch of a Gopher path, and that information is easily accessed by anyone who finds that end-branch via browsing using Gopher-proper or via Veronica.

Thus, an electronic journal that is distributed for free might have its archive of articles available at a Gopher end-branch.

Similarly the AAUP is actively tying together individual Gopher files of the catalogs of the University Presses, so that a single Gopher tunnel structure will reveal the distributed comprehensiveness of scholarly title output by these presses.

4. WAIS

As mentioned above, the Wide-Area Information Server is a true, natural language tool for searching for and retrieving textual documents. The documents may be located at diverse Internet hosts, but the index that is searched will be in one location. WAIS is the most "pure" information retrieval system generally available to Internet users.

Using WAIS, a publisher might index the full text of a scholarly journal, a collection of articles from diverse sources (assuming copyright problems were taken care of), or internal databases to be shared among workers at several locations.

5. World Wide Web

WWW represents the most recent information management "paradigm shift" instituted on the Internet. Although difficult to explain without the benefit of actual examples, WWW provides hypertext linking across the Internet.

Thus, an author might create a document that refers in a footnote to another document. That footnote in WWW is actually a hypertext link to the address (location) of the referenced document on some other host somewhere else in the world. By tabbing to and pressing the carriage return on (or in some implementations clicking on with a mouse) the hypertext reference, access to the remote document is actually achieved and the document displayed for immediate review. WWW works among textual material, graphic material, full-motion video, or virtually anything that may be displayed on a PC screen or played through a PC's speakers.

Douglas Engelbart and other architects of the original ARPANET considered many of these potential tools and linkages in the 1960s.⁽¹⁾ That it has taken over 30 years to produce working implementations of Doug's theories (to be sure, he actually *demonstrated* linking capacities and hypertextual referents in his *Augment* architecture) may be a comfort to some. That in thirty short years, a variety of inexpensive working tools have been widely distributed among thousands of users may be an inspiration to others.

6. Mosaic

Mosaic provides publication-quality appearance for WWW documents. Using SGML-like tags, or interpreting text format on the fly, Mosaic presents a high-quality document image in a graphical user interface environment (e.g., Windows, Mac, XWindows, etc.). Mosaic and WWW together make publishing on the Internet a reality.

A publisher implementing a Mosaic interpretation would not be restricted to low-resolution images. One could imagine many applications in which high resolution images were critical, for example, in medicine, astronomy, radiology, or astro-

physics. Such disciplines could benefit from an implementation that permitted scanning low-resolution images and selective identification of images to download in full color and exquisite resolution.

B. ACCESS AND NAVIGATION

While many of the advantages of Internet are yet dreams, access is a very tangible advantage all users currently enjoy. The Internet has opened up access to vast libraries of information, computing power, and user communities.

There are no directories that show you how to get around the Internet, but it is estimated that 5,000 discussion groups and 2,500 electronic newsletters exist in "cyberspace" over the Internet. Research shows that almost 100% of all scholars in the hard sciences and a majority of scholars in the other disciplines now use Internet E-Mail as a primary communication link with colleagues.

Almost all major archives—the Library of Congress, the British Library, the Bibliotheque de France—are all racing to make electronic archives of their literature accessible over Internet.

Hardware restrictions (modem speed, memory capacity, etc.) and software protocols (search systems and user-unfriendliness) currently limit many users of the Internet. But with hardware prices dropping and software becoming more sophisticated and user-friendly, it is merely a matter of time before these are no longer barriers of access for many presently disenfranchised users. Cellular technology promises to make modems accessible anywhere; and hardware becomes smaller and more ergonomic, bringing us ever closer to the day we can purchase a computer you can curl up in bed with to read.

And if not that, then certainly computers which can provide a more level playing field for the differently abled. Today, computers support Braille readers, joystick peripherals, and a variety of devices which bring to an entirely new market the riches produced by scholars and researchers.

C. IMMEDIACY

Discussion about the immediacy and responsiveness of the Internet has already been sprinkled through the text. An important question for professional and scholarly publishers is "Will the Internet shorten my publication cycle?" How much faster is networked distribution when peer review

and author revision cycles are taken into consideration? Probably not much. While the actual communications process can be accelerated, the intellectual work involved in moving a manuscript from one stage of development to another is only marginally influenced by the ability to get the product of one's efforts more quickly to the next person in line.

At present there has been a slight decrease in the speed at which manuscripts are reviewed and revised via electronic means. Reviewers and authors adapting to new systems where conversions must take place and new ways to edit their materials must be learned.

Yet some individuals prefer to print out the documents and make changes to the printed version.

As long as reviewers and authors are given set deadlines to return their materials, the speed of publication will not be affected, except perhaps by a few days.

On the other hand, an increasing number of editors are engaged in online editing, simultaneously identifying and keying needed changes. Such work is especially efficient if the resulting files can be sent directly to an imagesetter for typesetting, or onto a file server for eventual retrieval and redistribution.

D. COST

Costs can be measured in a variety of ways and in almost every case, cost/benefit ratios as applied to distributed networks favor the network over alternative forms of activity.

- Electronic storage is fast displacing paper as the most cost-effective storage mechanism.
- Digital archiving has obvious advantages over costly library stack shelving.
- File maintenance and "refresh" (needed to avoid bit drop over time or outdated media) is far less costly than microfilming aging books.
- Image storage permits enhancement technologies to improve the legibility of manuscript images, making digital storage more cost-effective than any other means of replicating such collections.
- Transmission costs are lower than any competing technology (phone, FAX, post, or ground transport).
- Sharing resources across a network, such as printers, powerful computers, and specialized

peripherals, provides enormous economies of scale for institutions of all sizes, and through them provides capabilities for individuals they could otherwise never afford.

Publishers who are reluctant to go to the trouble of learning the plethora of new information required for entry onto the Internet might do well to consider these advantages. They are real, and they are compelling.

PRIMARY DISADVANTAGES

of distributing information over the INTERNET

The following list of problems is not intended as a justification to "do nothing." In fact, it is in the identification of deficiencies that solutions can be found. The more explicitly publishers articulate their concerns about distributed network publishing, the more likely will the solutions be found.

It is difficult to categorize these problems, since any deficiency noted might or might not play a role in any given strategy a publisher might employ to exploit the Internet. For some, visual display will be critical to the effective presentation of specialized material. For others who choose to deliver information in plain-vanilla ASCII format, screen display issues will be inconsequential, but security might be a more serious problem.

A. POOR SCREEN DISPLAY QUALITY

The biggest reality with respect to screen display is the wide variety of terminals network users have at their disposal. For most users, character-generating devices are the lowest common denominator. These, for some time, are likely to be the majority of the types of hardware by which users access Internet information. A publisher who does not support this fundamental capability runs the risk of isolating a significant potential market. But providing support for both character-generating and graphics-capable video screens is something of an organizational (if not technical) challenge.

Edward Tufte (in *The Visual Display of Quantitative Information*) presents eloquent evidence of the high information-carrying capacity of paper, as compared with even the highest-resolution computer monitor. For specialized and detailed information, for material displayed in small point sizes of type, or for tabular or statistical material, these limitations can be significant.

B. INFORMATION QUALITY CONTROL

One appeal of the Internet is the ease with which vast quantities of raw material can be made available to equally vast numbers of potential

readers or users. This is equally its disadvantage. John Perry Barlow for the *Electronic Frontier Foundation* (and lyricist of *The Grateful Dead*) expressed it well in a speech made in December 1992 to the intelligence community (read: CIA) on network security:

"There is an enormous amount of information on the Net. But the signal-to-noise ratio...is terrible. You folks [the CIA] have some expertise in an important function: sorting out that which is relevant from the huge spray of data that is coming at everyone. That is an important problem that is largely overlooked...so far the software solutions to it don't strike me as being much good. ..[T]he difference between data and information is meaning, something machines know little of....you must pass [the data] through a human mind. There is also a question of authority, reliability, and bias."

This certification ("authority, reliability") process is what publishers manage every day. The post-publication filtering or sorting process has traditionally been the role of a combination of the information specialist and the end user. These distinctive roles are likely to continue, but as electronic publishing becomes more sophisticated, publishers will have to invest more human effort in the intellectual organization and electronic presentation of their certified information if the material is to make its way optimally through screens, filters, and human intervention to the readers for whom it will be of most benefit.

In paper, human eyes and the human brain can compensate for typos and layout errors and can identify visually unlabeled information (e.g., that "Smith" is the name of the author). In an electronic environment, everything must be structured in order to be retrieved. Links should be made between related concepts or documents. Publishers will have to think harder about how their publications might be used and by whom and build into their material the necessary access tools to make it both retrievable as such and filterable so that it goes to the proper users.

Successful electronic narrowcasting may be the ultimate goal of most publishers. As one scientist described his ideal information system: "I want to receive all of the information I need and only the information I need."

C. HIDDEN REAL COSTS AND USER EXPECTATIONS

Internet users, particularly academic users, have perceptions about the Internet which handicap publishers. The Internet seems free to the academic user. The costs of construction and use are borne by governments and institutions and are rarely charged directly to the user. Because the Internet seems free and it appears to be ubiquitous, user expectations about the ability to distribute information freely and widely over the Net are high. Publishers will find with increasing regularity that they will be asked to share their right of distribution with authors who want to put their article, chapter, or book on a file server for network access, as well.

Users and librarians also have expectations about the relative cost of paper and electronic versions of the same material. Their idea is that the simple savings of paper, printing, and mailing should translate directly into a lower electronic price. In fact, at a minimum, parallel electronic editions of works that exist also in paper have their own development and handling costs. To the degree that both a paper and an electronic version are created, the total costs are higher and must be borne by all delivery media.

There are other hidden and visible costs. These include the intellectual efforts to improve the value and retrievability of the electronic material. Journal abstracts, for example, may need to be more standardized and better written. Book chapters should be subdivided into logical, searchable sections. SGML tagging will need to be done. The marketing and sales infrastructure for most electronic products is much more extensive and expensive than for a comparable print product. The same is true for customer support after the sale. One rarely has a complaint from a book buyer that the book would not "boot."

Libraries and readers do not easily accept the economic realities of electronic publishing, and expect resources/information to be free or at least less-expensive (or, like hardware and software, decreasing in cost from year to year). This is a disadvantage but not an insurmountable handicap. It places an obligation on publishers to add greater perceived value.

D. COPYRIGHT, SECURITY, AND RELATED ISSUES

One of the most volatile of topics surrounding electronic publishing is the issue of how copyright will apply in the digital environment.

In the discussion above, copyright has been mentioned in the abstract, as either supported or challenged by technical and software which permit various ways to handle the information.

There are fundamental challenges to publishers from sectors of the community who feel it would be best not to transfer rights to publishers at all. This strikes at the heart of the publishing enterprise, and some consideration deserves thought.

Proposals about authors retaining copyright are usually couched in response to a perception that authors strive to create works of intellectual merit; publishers simply obtain the rights to these works and sell them back to the authors (or their institutions) at inflated prices.

This bald characterisation, for authors, falls under the rubric of "Buying Back [Their Own] Research" from publishers. But research results generated at a site (be it a university, research institute, or corporate lab) are typically issued as Technical Reports.

Most institutions already manage their own Technical Reports. So what they are buying from the publisher is actually something different; or they might simply drop the publishing, since the research results are already theirs.

There is a widespread undervaluing of publishers' contributions to the exchange of scholarly information. Most of the models being considered for new arrangements in the management of copyright, or at least their proponents will acknowledge that publishers add value. But it is questionable how deep this acknowledgment goes. There seems to be a major underlying assumption of the Primacy of the Researcher in knowledge creation coupled with the notion that publishers perform a subsidiary role in 'prettying up' the content, adding value in a minor way entirely out of proportion to the prices charged. This view also treats libraries and computing centers as mere service functions.

In an earlier section, this naive understanding of publishers' added value may have been dispelled. Considering an alternate view of knowledge cre-

ation (one that makes more and more sense in the electronic environment) reveals it to be a *process* that merely *includes* the researcher's work, as it does the additional contributions of other participants in the process.

Any number of other stakeholders aid in support of the creative process: institutions that provide facilities and resources (like libraries and laboratories); organizations that bring together minds with similar interests to sharpen focus and stimulate advance; ongoing local tradition or schools of thought to situate and contextualize an individual researcher in his work and the moneys provided for the research to take place, whether collected through tax dollars, private donations, or salaries. (It is considerations like these that lend substance to the claim that faculty writings should be treated as works-for-hire.)

All these functions are provided by the public, the government, the institution, the publisher, the editor, the library, and other scholars in the process of knowledge creation. From such a sociological perspective, the individual researcher (or author) can just as easily be viewed as a service provider in the chain of knowledge creation, giving back the results of intellectual investigations for others to review, refine, re-present, organize, index, preserve, etc. The scholar may be accorded a privileged position in this elaborate and costly process of knowledge creation, if one wishes. Certainly, the anonymous work is not a strong feature of our culture.

E. PRIVACY

There is great sense that the network will provide the most valuable marketing research available: To know user search patterns might enable the automatic delivery of specific targeted advertisements between searches, much as the electronic grocery check-out prints up specialized coupons at the supermarket based on the products a shopper purchases.

Librarians legitimately fear the misuse of information collected from search profiles or from records of what information individual users may have accessed.

When analyzing user behavior, many questions need to be asked. How should one record an actual search? Perhaps the dialog that ensues after the search is more valuable. Perhaps all the notes prior

to the paper are more useful? What qualifies as the peer-reviewed portion? How much backtracking is taking place? Does that usage pattern have any significance in future publishing investments? Are certain articles more commonly read than others? What conclusions might be drawn? How much of such information can be shared?

F. BILLING AND TRANSACTION INSTRUMENTATION

The absence of monitoring and billing instrumentation is a fundamental problem. Until methods are developed by which users can be billed for access to information in incremental and flexible ways, there will be a natural hesitancy for serious publishers to enter the Internet market with intellectual products that may have resulted from significant investments of time, energy and money.

Enforcing copyright depends on an ability to make payments to the licensed owner of a work. Whether one should track fractions of a penny for each use, or generalize to a site license schema, the present absence of tools with which to monitor and measure use, and with which to make automatic monetary exchanges for permissions is a major barrier to professional publishers.

G. ARCHIVING

The archiving issue is not obvious. Even the Library of Congress discards hard copy issues after a number of years. Some think that publishers should be responsible for archiving their own materials, or at least work in conjunction with an assigned library or other organizations to house their historical work. Individual libraries cannot archive all materials, but special libraries should work with organizations or societies to archive older works in specialized disciplines.

Archiving electronic works will be more difficult, especially the electronic newsletters, serials, and bulletin boards containing comments on a variety of issues. In such cases, should archiving responsibilities reside with the database publisher or producer? This would be an area where a library or other group could work in conjunction with the database publisher and/or producer to assure that archival copies are maintained?

Archival considerations for electronic files actually involve two components: *organizational* and *physical*.

On the organizational level, there is a social requirement that materials that are to be archived be dealt with in a relatively stable ongoing environment. In our society, libraries (both municipal public and institutional private ones) have accepted that role, becoming involved in preservation, restoration, rebinding, and a variety of other specializations by which printed books are preserved over long periods of time (at least beyond several human lifespans) for the good of society.

It is attractive to think of a publisher mounting and being responsible for the archival depository of its product line. That way, as revisions come in, files can be updated immediately, and the available "edition" would always be the most current. As the bisynchronous nature of the network becomes more manifest in publishing, there might be annotations to files that ought to be kept along with them. Hyperlinks will attach to the files or emanate from them. All these aspects call for *control* and *stability*.

One of the important questions publishers ought to solve in cooperation with their library colleagues is just who is to be responsible for the archiving of published works, and at what point may they be considered "frozen" in an edition that is truly of archival quality. It is possible there could be two stages in the process: The publisher could manage its own archival repositories for a time, and assemble whatever annotations and commentary deserves to be attached to a file. After a suitable period of time the files might be handed over to the library community to be formally *archived*, while the dynamic file remains at the publishers for the ongoing matter of new commentary, citation, or changes.

Whatever arrangement needs to be worked out, there is no consensus on the matter, and only fuzzy understanding about why "archiving" is a notion and that means different things to librarians, publishers, and Internauts.

The second component in the discussion involves file integrity at a physical level. Are there redundant copies resident on different host computers in case of a disk crash? Is someone taking responsibility for refreshing electronic media periodically to eliminate the risk of naturally-occurring bit-drop? Is there a function that routinely inspects files and compares them with some original master file? When are such precautions important and when are they merely expensive?

Additional issues are the question of backup copies on other-than-electronic media; the proper archiving of files containing motion sequences, or ones with many hyperlinks to external documents that may, themselves, be ephemeral and in danger of disappearing.

Symposia might be organized to evaluate all these concerns from the perspectives of the larger scholarly community, and the smaller community of professional and scholarly publishers before the PSP community makes *ad hoc* decisions that could prove troublesome to other constituents.

H. TENURE AND MERIT

Tenure and merit are usually matters quite beyond the purview of the PSP community. The topic is raised here as a *disadvantage* of the Internet environment to permit the identification of the social acceptability of exclusively electronic publication.

The situation resembles the economic spiral which was discussed at the very beginning of this briefing paper.

- Electronic publishing on the Internet lacks the formal rigor of publication on paper.
- The absence of the guarantors of objective peer-review and professional editorial processing (i.e., PSP members and their allies) has left a niche open for all kinds of entrepreneurial publishing, some of which approximates the standards of print; some of which may exceed the standards of print; and some of which is decidedly inferior to the standards of print.
- In this environment, tenure and review committees are reluctant to give credit for exclusively electronic publication because they do not yet possess an apparent or implied litmus test for the value of publications appearing on the Internet. Being a naturally conservative function, these committees choose not to recognize electronic publication as meritorious of scholarly advancement;
- Watching such decisions, publishers conclude that the electronic environment is far too risky a venture with which to become involved.

Until electronic publications gain the respectability of their print predecessors, the evolution of electronic publication as a first option for many authors will be greatly inhibited.

The solutions to this dilemma may be several:

- Simultaneous paper and electronic publication;
- Delayed paper publication;
- Endorsement by a Society or credible "imprint giver" which endorsement could be recorded as a distinction for certain types of publications;
- Formal acknowledgement of a difference between peer-reviewed and professionally handled electronic publications, and those which are merely self-monitored.

The point is—as it is with so many considerations involving Internet publishing—that several options exist, but few will be explored unless the option of publishing on the Internet is recognized, not as a risk to be taken by individual professional and scholarly publishers, but as a necessity mandated by the broad coalition of those involved with scholarly communication.

DRAFT

PART IV — MIGRATING TO THE ELECTRONIC MILIEU

The challenge of this section is to provide a roadmap on "how to get started," and it is not taken lightly by the authors of the papers comprising this text.

It is, however, difficult to chart a path in the absence of leaders from whom one can learn the way to avoid litigation or bankruptcy. This is one of the reasons for the rise in cooperatives and consortia: to diminish the risk and increase the likelihood of expert opinion coming from colleagues external to the publishing profession.

Not everyone, of course, has the resources or the luxury to involve themselves with external partners as have those who are involved with the major prototype experiments listed in the Appendix. There is, however, a general asset that all professional and scholarly publishers possess, and that is their *authors*.

One of the sad results of professionalism in the scholarly communications industry is that the professionalization of editorial and publishing activities tends to segregate the publisher from the aca-

ademic or research scholar. The academic is already under pressure to specialize within academic disciplines, and the distance between the work of the scholar and the work of the publisher can be huge.

Professional and scholarly publishers emerged from the fecund environment of stimulating thought and discourse carried on around specific subject areas. Because that discourse was so exciting and productive it needed to be shared. It was because it was so lively that it needed to be "pulled off" (as steam is from a pressure-cooker) the dialog uninhibited by the practicalities of peer review, editorial processing, file preparation, manufacturing, and marketing.

Most authors are more advanced in their knowledge about the network than are their publishers. This has unnerved many publishers, and disappointed many authors, but it can be turned into an advantage as well.

It is clear that only a small (but growing) proportion of traditional publishing houses are connected to the Internet, and are comfortable in using its tools.

- The thrust towards improved communications with authors (and the author's friends) may be all that is needed to justify the installation of an Internet-capable E-Mail connection.
- Authors may have interesting electronic projects for which they need expert advice.
- There may be an opportunity to finance an innovative network project which might not gen-

erate any revenues, but which could begin to recreate the loyal relationships which many publishers (as well as authors) claim have diminished over the years.

- Authors with introductions to university administrators could bring together the professional publisher and the professional administrator, from which might arise any number of (licensing, sub-contracting, projects with economic-guarantees) ways in which a creative dialog among *partners* rather than *vendors*, could break many of the barriers, or reduce many of the fears of experimentation.

If these suggestions seem too optimistic, too time-consuming, or too unlikely of beneficial return, there are more practical things publishers could begin to do to prepare for taking up a role in the expanding distributed network environment.

- Establish a liaison with the network managers (perhaps IETF) concerned with standards, informing them about the kinds of information management apprehensions publishers may have, and to work with them at resolving such limitations;
- Establish a modest (cash) award for the most innovative electronic publication effort undertaken by an author during the past year. (Perhaps one corollary of this suggestion might be that a publisher must make the nomination and would not be inhibited from providing advice to an author. But the publisher may not *have to be* the electronic publisher of record.)
- Decide that all official correspondence from committees and administration be distributed electronically and mounted on a gopher. This would go a long way towards reducing Association costs, and—at the same time—provide a practical benefit for those staff members involved in committee work, program planning, etc.
- Establish a ListServ by individual publishers for their authors, within which to discuss matters concerning electronic publishing projects. This would provide the publisher with a two-way communication with the community it serves.
- Identify a staff member to join the ListServes in specific subject areas important to that publisher, and produce a regular report concerning the most interesting information, leads about authors, hints about threads of investigation,

and marketing possibilities deriving from such a source.

- Publishers ought to learn from the example of the academic networks from which the Internet has taken its greatest growth leaps. Providing accounts and E-Mail access is neither a luxury nor a line-item budget. Every employee using a PC and capable of sharing a modem on the Net should be assigned an account number, provided with instructions, and encouraged to use the Net for all manner of communications, research, and file management. The result of this policy will be an enormous cadre of available workers familiar with the Net.

Just "Do it" is a contemporary call to which too few established businesses pay enough heed. In the case of exploring the Internet, it may be good advice, and it might provide an interesting respite from the harries of an average publisher's work day. Many Internauts do things which are "fun" rather than "productive." It is, perhaps, another reason for the sometimes ill-hidden cynicism with which traditional publishers regard network activities. You can hear them mutter: "They wouldn't do that if they were trying to keep their businesses solvent!"

Perhaps the Internauts don't know how to run a business on the Internet for the same reasons publishers don't. Betting on which could do so given the same start, most people would bet the publisher would succeed faster than the non publisher. On the other hand, perhaps some of the enthusiastic enjoyment that appears to be characteristic of present Internet activities might reinvigorate pockets of our industry that might benefit from it.

When contemplating a move to network publishing, it is important to understand the two rather fundamentally different ways of preparing files and formatting them for the electronic environment.

FILE PREPARATION AND FORMATTING

Publishers who are considering whether and how to get into electronic publishing have even more basic considerations. They have always had to deal with the manufacturing and purchasing aspects of their business. Most of the time, these functions were outsourced to suppliers and vendors. But for many professional and scholarly publishers, the emergence of powerful desktop publishing software

programs that could run on relatively inexpensive personal computers appeared to be an attractive way of reducing out-of-pocket costs, regaining control over manuscript processing, and permitting economies of both labor and time.

This operation is also a value provided by the publishers to their audience: the preparation of files suitable for public display and distribution.

Once again, however, local decisions which seem to have self-evident justification may sometimes run counter to a long-term strategic need. So it is worth trying to understand better the ramifications and options available for file preparation and formatting as a precursor for engaging in the complexities of Internet publishing.

Publishers should be aware that after the acquisitions, editorial, and marketing functions have been accomplished, there remains the task of preparing files suitable for manufacturing (or for electronic dissemination). Two fundamentally different approaches to the preparation of files for publication have developed over the past five years. In simplified terms, one of these might be called a *page-oriented* approach and the other a *database-oriented* approach. Each approach has its own set of technologies; each has its own cost structure; and each anticipates different reader/user needs.

Clearly, the decision to opt for one or the other method of file preparation is important, perhaps more important than many publishers realize.

1. Page-Orientation

The page-oriented approach preserves the greatest continuity with publications as they are now known and with the publishing business as it is now understood. In its pure form, this approach will prepare publications in traditional mode for print. Then digital pictures of the printed publication will be made using scanning technology that produces raster images of the pages for bitmapped display. To form a "wrapper" for the publication that might approximate the physical binding of paper, the files containing the page images must be organized such that they can be accessed sequentially on screen or printed out in linear sequence from front to back covers.

Since page images cannot be searched, a separate index or table of contents needs to be prepared so that documents can be identified and retrieved as complete entities.

There are some advantages to the page-oriented approach:

- Documents in this form are not editable. Therefore the integrity of published contents are more easily assured.
- Documents look the same with regard to layout and typography.
- Nontextual material such as equations, tables, figures, and artwork may all be captured in the same consistent manner, and can more easily be rendered again for print and display.
- Older publications are as easily captured as current ones, and the costs of capture are not especially high, probably less than \$1.00 per page.
- This approach fits fairly easily into an existing print production flow, allowing final access or printing to take place locally after distribution from a central production site.
- Perhaps the most compelling argument in favor of page-description approaches is not the ubiquity of applications programs that support PostScript (a primary language facilitating page-oriented makeup), but that people are used to journals with pages formatted in a particular familiar style and sequence.
- The page-oriented approach maintains the familiar look and feel of existing publications and requires the least adaptation in the knowledge acquisition habits of the reader. If printed versions of the publications decline, there are cost savings to offset the additional manufacturing expense. More often, large collections of page images could lead to new product offerings rather than becoming a competing alternative to print.
- Desktop publishing tools can be used to support page-oriented electronic publishing precisely because they were designed with the final form printed page in mind. This makes it relatively easy for even an inexperienced operator to construct pages that have basic and attractive appearance, and can be saved as files which can be manipulated by any of a variety of readily available software applications.

In the page-oriented approach, communication networks, such as the Internet, are seen as transmission vehicles of published works, rather than as actual publication media.

There exist disadvantages to page-oriented electronic publishing:

- It fails to take full advantage of new value-adding possibilities inherent in the digital environment, (although it may minimize the risks). This may be the biggest disadvantage.
- Page image files are very large. Hence they require a lot of space. The more faithfully they represent the original page, the higher the scanning resolution must be, and the more bits of data required to represent every picture element or pixel.
- Given a limited bandwidth, transmission time can be very slow, relative to files that are symbolically encoded. There are various methods of making these files smaller, called compression algorithms (e.g., CCITT Group III, CCITT Group IV (FAX), GIF, PCX, PICT, JPEG for rich color and photographs, MPEG for digital video), but they will still be much larger than similarly compressed text files and take longer to transmit. There are a number of representation formats for raster images to which compression can be applied. GIF (Graphics Interface Format) is a popular low level representation as well as a compression technique. TIFF (Tagged Image File Format), is a more formalized and accurate representation to which Group IV compression is most commonly applied.
- Page-oriented files do not readily lend themselves to the database approach because they produce files that intertwine presentation information and processing instructions with data. Moreover, the file formats produced are platform- and application-dependent.

Recognizing these shortcomings of page-oriented electronic publishing tools, the industry that supplies them has been moving to add value to their offerings to improve both the device independence and the usability of their final form, electronic pages.

PostScript® is a page description language that can represent text as a sequence of characters, graphics as a series of operations that rebuild the picture, and both binary and continuous tone raster images. PostScript is the output of a composition and layout program. Although widespread, there are many non-PostScript printers that cannot render PostScript files, and there are many platforms that

do not have viewing software that can display PostScript files on screen.

In an attempt to make PostScript files universally printable and viewable, Adobe (the developers and publishers of PostScript) has developed *Acrobat*®.

Inside *Acrobat* there is something called the Distiller that can transform PostScript into PDF (Portable Document Format) files. Each platform requires its own fairly inexpensive viewer to display PDF files on screen and to print them to any printer.

Another component of *Acrobat* called *SuperATM*, also available unbundled, provides a font substitution mechanism such that the recipient of the document does not require the fonts the originator had when creating the document.

Acrobat thus produces displayable and printable files that are smaller, faster, and more device-independent than PostScript. In addition, while not truly a revisable or editable format, annotating and linking documents is supported, as are certain types of searching.

Thus *Acrobat* is helping to facilitate a significant enhancement to desktop publishing based on page-description approaches. It is still oriented toward the printed page, but it goes some distance toward enhancing document interchange across platforms and providing additional functionalities useful to the scholar.

Both PostScript and *Acrobat* have also provided a reliable transportable file, which can easily be transmitted across the Internet without loss of information. This cannot be said of proprietary typesetting files which contain large numbers of non-ASCII control codes.

Together with TeX, (which is a typesetting language of particular use to mathematicians, producing files that are input to a composition and layout program), PostScript and *Acrobat* facilitate a different re-organization of publishing.

These are the tools of self-publishing authors, and self-reviewing and self-credentialing networked communities in which publishers may play no part, or perhaps a very reduced part, and money may not directly enter into the scholarly exchange.

2. Database Mechanisms

The database-oriented approach to electronic publishing is a much more radical departure from traditional methods. It requires different expensive technology. It requires different skills to operate, on top of the traditional editorial, production, and design skills typically found in a publishing house. This means additional costs in both people and training. It also means substantial change in the workplace and the anxiety and resistance that can engender. It is, however, the high ground of transportable files, adopted by government, international standards organizations, the military, and any number of publishers of optical or CD-ROM data.

The database-oriented approach is facilitated by SGML (Standard Generalized Markup Language). This protocol liberates the *structure* and *content* of a publication from its presentation. This enables a publication and all its elements to be treated as *objects*. These objects may be manipulated in many ways that page-images cannot.

SGML is an international standard that specifies the rules for developing applications of electronic manuscript markup. When one applies the standard a machine-readable Document Type Definition, or DTD, results that describes the elements that may occur in this type of document. It also specifies the order in which they may occur and what attributes they may have.

Preparing a SGML-conforming file yields a document "instance" that conforms to the DTD. An SGML file contains only ASCII characters. As a result SGML files may be exchanged among all platforms that accept ASCII. The DTD is passed along with the document instantiation.

The receiver who has an SGML parser can read the DTD and verify that the document conforms to it. There are some authoring tools available that hide the intricacies of the complex markup from the author behind more friendly displays, but they are not yet nearly as widespread as page-oriented word processors.

Well-known examples are SoftQuad's *Author/Editor* and Arbortext's *Adept* series of tools. *WordPerfect* has also developed a companion product that provides some support for the creation of SGML conforming documents, though only on export.

The desktop world in general is becoming more aware of SGML and is moving to create built-in support for this scheme, especially since its military counterpart, CALS, has been mandated as the mechanism of transmission for vendors and contractors for submitting proposals to the military.

The strategy of adopting a database-oriented approach is based on the premise that, in the long run, the greatest advantage of digital media for scholarly research and publication comes not from transmission of page images for local display or printing, but from the generalizable standard mechanism by which storage, reuse, and platform independence can be guaranteed.

This approach predicts that scholarship will benefit from the ability to search, browse, annotate, hyperlink, and embed non-textual objects like voice, animation, and full-motion into files. It permits the expectation that files can be further processed or otherwise interacted with in an unknown future; one that can be sure to be backwards-compatible with a ubiquitous and generalized standard. SGML permits unlimited manipulation of publication contents, either for searching or for display in various formats. It also assumes that the scholar wants a common set of tools that will interact with ANY publisher's information.

The database approach yields a number of advantages:

- It is a standard approach to tagging manuscripts can be taught throughout the industry, developing a cadre of skilled labor and promoting consistency;
- Files are immune from the vagaries of software development and hardware upgrades would relieve considerable pressure from the publishing community;
- The database approach can provide the publisher with new print revenue opportunities in the short run. (Managing an electronic resource of a publisher's entire holdings can facilitate re-packaging and re-selling various views of that body of literature for different audiences at different times.)

There are negatives, as well, to consider:

- The SGML-tagging approach represents a professional, not a popular standard: It will be more difficult to convince authors to conform to the discipline required for SGML tagging, es-

pecially since few authoring tools are presently available to support such a request;

- The single greatest drawback to the database-oriented approach to electronic publishing is the difficulty of rendering to page or screen.
- If presentation format has once been separated from content, the technologies available to represent the content are few and not that easy to use.
- While SGML is an established international standard, the DSSSL standard, with its associated FOSIs (File Output Specification Instance), is not yet approved and implemented.
- The existing tools for SGML tagging are somewhat cumbersome, and require a special understanding of the construction of manuscripts which is not widely understood by the general public.
- The setup investment and the training support required for SGML is of a higher order of magnitude (given its long-term advantages) than its page-description applications (given their short-term immediacy).

The economics pose the fundamental quandary publishers face: Those interested in long-term benefits will see the advantage of adopting an SGML approach. Those interested in short-term and immediate benefits will find the object-oriented page layout approach far more appealing.

In either case, personnel will be needed to master the formatting tools that are ultimately adopted, and strategists will have to develop business plans that may differ depending on the practical decision that has been made about the generation of publication-quality files by either approach.

The organization of knowledge is something which the publishing industry both preserves and defines, albeit in preconscious ways. Chaos reigns on the networks; knowledge and information sources are not well-organized; great expertise is required to navigate in this information space. Publishing skills are greatly needed to establish the networks as a truly viable scholarly medium. Publishers are in an excellent position to explicate how value is added to raw information and how this is to be translated on a functional plane to the digital medium. To do so, they will need a better appreciation for the capabilities and limitations of the networked environment.

PART V — CONCLUSIONS

Early in this paper the observation was made that scholarly communicators are fascinated by the potential of the Internet for enhancing and facilitating their work.

The editor of these essays can report that one of the most difficult (maybe least successful) challenges was to tone down the language of approbation that surrounded any discussion of Internet potentialities for professional and scholarly publishers. This enthusiasm is out of proportion with the actual business implementations by the PSP community, so the difference is notable.

If the collected essays are reflective of a wider opinion among professional and scholarly publishers, then the time at which professional publishing actually comes to the Net is far closer than any of us may be predicting.

With apologies, following is a somewhat simplistic preliminary effort to look at conventional publishing activities, and to begin to chart them in a manner permitting more formal evaluation. This first draft attempt is provided in the hopes it can be improved by local and specific modification to develop some long-range planning for those publishers inclined to do engage in such discussion.

Activity	Traditional Elements	Networked Advantages	Networked Disadvantages
Acquisition:	Identifying Authors Travelers Lunches	The network is bisynchronous: communication goes in both directions. The Internet provides excellent possibilities for identifying authors, conducting preliminary negotiations, and receiving preliminary outlines and MSS.	Virtual Lunches are less satisfying, even though they are less fattening.

	Activity	Traditional Elements	Networked Advantages	Networked Disadvantages
Editorial:	MS Review Process	Mailing, Tracking, Followup, Trafficking.	The network can facilitate the processing, but, as was pointed out in the narrative, the intellectual labor is barely affected by speed of communication. Various tools can help the staff effort involved in tracking replies and organizing the traffic associated with the review process.	The network efficiently connects with those who have E-Mail addresses. Since not all PSP reviewers are likely to exist on the net, procedures will have to be developed to reach those non-networked individuals and integrate those transactions with the ones being tracked electronically.
	Copy Editing	Reading, Correcting, Marking up.	Since the natural "language" of the Internet is ASCII, files which can be transmitted in this way go some distance to being "regularized" for editorial processing. ASCII files can be imported into either page layout or database format (see Part IV).	ASCII can't support mathematics or special diacritics so is of limited use.
	Citation Verification	Checking, Looking up, Correcting	The proliferation of OPACs on the net permits a vast array of bibliographic resources with which to make such work easier.	
	Version Control	Making sure what's what and where it is.		No change. Version control is exceedingly important in an electronic environment because it is exceedingly easy to make copies, send duplicate files, become confused about naming conventions, etc. As multiple authoring tools arise, the problem of version control will accelerate. No matter what software tools may be developed, this is largely a detail job for an intelligent being.
Design:	Illustration	Much design is presently electronic, using specialized graphics or layout programs preparatory for submission to manufacturing.	No change. Binary files can be transmitted across the Internet. An increasing number of visualization tools exist for viewing files in various graphics formats and compression schemes.	
	Markup	Sometimes manual, sometimes automated, markup can be accomplished for two purposes: tagging elements and tagging typographic information.	No change. Files could, conceivably, be delivered to typesetters, or they could be sent to remote sites for further processing. The emergence of network printing devices permits files to be printed remotely, on specialized equipment, or supplied directly to a customer.	
Production:	Technical Knowledge	"Translating" plain English to various vendors who speak a variety of "technical" jargons.	No change. Many of the computer-related terms production people have learned over the past decade during which typesetters and manufacturing plants became highly computerized are applicable to networks. The "burnt fingers" are also applicable, since many of the things learned from computerized typesetting are directly transferable to distributed networks.	
	Preparation of CRC	MS Preparation, desktop publishing proof checking, final integration of art.	No change. No great advantage, either.	There may be a need for a different set of procedures for a paper-ready file, and one that is computer-ready for electronic publishing. At present, only one desktop publishing layout program supports network needs and hypertext, although the others are investigating enhancements.

	Activity	Traditional Elements	Networked Advantages	Networked Disadvantages
Marketing:	Purchasing	Negotiating Prices, Evaluating Bids, Letting Contracts.	No change.	Few established vendors are connected to the Internet or by E-Mail, FAX seeming to be the preferable communications choice.
	Audience "building"		The networks can-through the judicious use of ListSerts and BBSs, provide a mechanism for supporting informal discussions among a publishers' constituents.	Not comprehensive. Some markets are apt to have a critical mass on the Internet, while others may be sparsely connected. Will take investigation.
Advertising:	Advertising		The network's <i>reach</i> is limitless. The mechanisms by which to introduce advertising will surely come from commercial parties. A PSP implementation could be less disruptive; more beneficial. Might be well received.	The network culture is biased against advertising.
Warehouse:	Storage, Retrieval, and Handling	Space, rent, large inventory systems, heavy costs.	Far more efficient cost-wise. Kept in appropriate standardized formats, data and text files could be re-used and repackaged in a variety of ways.	
Fulfillment:	Subscription List-keeping, mailing, checking.		Functionally identical, but costs greatly reduced, and postal costs almost eliminated.	
Business:	Managing Royalties			No system for tracking usage or metering access. Variables of "scanning" "reading" and "copying" are distinctions that may carry proportional pricing. (see above)
	Accounting/Billing Generating Revenues			Lack of proven business models. Current favorites include: <ul style="list-style-type: none"> • Site License • Pay-per-view • Pay for connection • Controlled Subscription • Augmented print • Augmented electronic.

APPENDIX: CASE STUDIES

A. Journal Fulfillment And Document Delivery Pilots

Given the promise of the Internet as a mechanism for furthering the work in which professional and scholarly publishers are engaged, an impressive number of these publishers have undertaken experiments and pilot projects that are intended to lend experiential data about user behavior, practicalities of network transport, mechanics of billing, and other issues identified, above, as potential or real barriers. What is remarkable, is how far ahead these publishers seem to be compared with their usually cutting-edge trade brethren. It may be that trade publishers have been involved in conglomerate and joint ventures with the "infotainment" industry. But the likely explanation is the close alliance between the PSP community and their authors and readers (more so than is the case among other categories of publishers). Since these individuals are presently the most active users of the Internet for institutional, personal, and collaborative communications and information exchange, it is natural to assume their publishers would enter this new environment alongside their constituents. Not all the listings below are sponsored by professional or scholarly publishers, but all the prototypes involve PSP publications.

Document delivery has found a very warm and secure home on the Internet. Born just a few years after the first Xerox copier, document delivery has developed into a large business over the past two decades. However, its growth was somewhat restricted by the labor-intensive process of photocopying and mailing. The advent of electro-storage of article images, large capacity Internet connections, and sophisticated compression algorithms promise to make document delivery grow at an exponential rate.

All these hardware and software changes have been accompanied by an attitude change within the library community. This change was best characterized by Dick Rowe, while he was head of The Faxon Company, as a shift from "just in case" to "just in time" collection development.

Many libraries currently offer (or plan to offer in the near future) a seamless access system to documents over a network. Anne McKee, Bibliographic Services Librarian at Arizona State University's Fletcher Library, describes the process at her library:

"Through our shared CARL catalog, the users can note exactly what each campus has to offer. If the material they need is at the other campus, they bring the CARL print-out to our

reference desk, and the request is transmitted to the other campus via a fax modem. Through special software, our computer "speaks" to Tempe's fax machine. This provides seamless access to the entire Arizona State University Libraries collection with an average turnaround of 48 hours."

(*The Serials Librarian*, v.23, n. 3/4, 1993, pp 209-210.)

This is a fairly standard system at most libraries and it is only a matter of time before this system is fully integrated on an Internet platform and access to information is almost instantaneous and available to any user with a password. With this as a background, it is possible to review the options available to libraries and researchers for document delivery that can tap into Internet.

Using Internet for document delivery is still in the evolutionary stages. There are currently three main uses of the Internet for document delivery:

1. As a means to transfer orders for articles to a traditional document delivery service (e.g. *InfoStore*, UMI's *Article Clearinghouse*, ISI's *Genuine Article*, and *UnCover*) which can be seamlessly linked to the library's existing public access catalog systems (as described above).
2. As the actual document delivery service itself (e.g., an Internet gateway via *Dialog America Online*, etc.). These services currently provide only full-text articles. (no graphics)
3. As a link-up to a cache of images that are transported over the campus network (e.g., Carnegie Mellon University, University of Michigan, etc.). *Ariel* software produced by RLG allows images to be sent and received over the Internet. However, both sender and receiver must have software for it to work.

The following services are examined in some detail for the benefit of interested readers.

RESEARCH LIBRARIES GROUP, INC.
1200 Villa Street
Mountain View, CA 94941-1100
(415) 691-2240

ARIEL

Ariel is a PC-based document transmission system for use by libraries to facilitate their interlibrary document delivery traffic. Article images are scanned by an originating library and are transmitted to a receiving library over the Internet (thus eliminating telephone charges). Both libraries must be running *Ariel* software on their respective workstations.

The image sent by the lending library in response to a request by a borrowing library is printed at the receiving library on a laser printer and delivered to the patron. At present, downloading or storing of articles is not feasible. RLG is reported to be working on a version of *Ariel* that would allow both on-screen display of images, as well as their storage for eventual reuse.

Ariel is receiving mixed reviews: some librarians are enthusiastic about the quality of the resulting image and the speed with which an image can be captured and transmitted; others cite problems (but these are principally related to hardware or software configurations and installation, and will likely be temporary in nature). Most libraries tend to view *Ariel* traffic as ordinary interlibrary "loan" therefore they expect to assess no copyright fee for the individual articles, other than what librarians would consider to be beyond "fair use." There is some evidence that this may be interpreted differently from library to library.

Dr. David Wood
Deputy Director
BRITISH LIBRARY DOCUMENT SUPPLY
CENTRE
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
United Kingdom
+ (0937) 54-60-32
+ (0937) 54-63-33

BLDSC

The British Library Document Supply Center is the world's largest traditional document supplier, with over 200,000 periodicals. Recently, a table of contents service was inaugurated, marketed by BLDSC and EBSCO (see below). The Table of Contents are loaded onto the Internet. Orders are taken over the Internet. Articles are faxed or mailed.

BLDSC is seeking publishers' agreements to conduct a one-year pilot project in which BLDSC would be allowed to scan and store journal articles. This would constitute a change from similar systems, because BLDSC would presumably engage in systematic scanning and storage; not just after an article has been requested.

In exchange, BLDSC is promising to supply user data at the article level. BLDSC fulfillment operations will continue via traditional means (post), but may include a trial of electronic delivery.

CISTI
Montreal Road Campus of NRC
Building M55
Ottawa, Ontario
Canada K1A 0S2
(613) 993-1585

CISTI

CISTI is largely a document delivery service whose orders are taken over the Internet.

CUPID
A Working Group of
The Coalition for Networked Information
21 Dupont Circle
Washington, D.C. 20036
(202) 986-5532

CUPID

CUPID (Consortium for University Printing and Information Distribution) involves of a number of prominent universities and Xerox, Canon and Kodak Corporations. CUPID's goal is to demonstrate the feasibility of high-quality distributed printing at remote sites. Such a capacity would support custom publishing, networked print-on-demand services, and even rare or hurt book preservation, all from one electronic master.

An example of how CUPID might function is the case in which a science journal has disappeared from the library. From a workstation, a professor accesses the journal across the Internet from a library which has a copy in its collection. The professor orders 150 copies to be printed at the university's copy shop for pick up within a few hours.

CUPID permits a wide implementation of billing and transaction analysis, so it can conform to many copyright and payment schedules.

DIALOG INFORMATION SERVICES, INC.
3460 Hillview Avenue
Palo Alto, CA 94304
(415) 858-3742

DIALOG

This major database information provider has had the most experience in online searching and information delivery. It is worth noting that Dialog currently provides a gateway into Internet, through which many of the Dialog databases can be accessed.

EBSCO SUBSCRIPTION SERVICES
P.O. Box 2453
Birmingham, AL 35202
(800) 633-4604

EBSCO

EBSCO will make the BLDSC (see above) table of contents database available to its customers via their proprietary network, EBSCO Net. Participating libraries can place orders for documents online; BLDSC will take care of fulfillment, either directly (post), or via EBSCO. Copyright fees will be paid following the recently announced BLDSC agreement to pay full royalties to U.S. publishers.

ELSEVIER SCIENCE PUBLISHERS
655 Avenue of the Americas
New York, NY 10010
(212) 989-5800

ELSEVIER SERVICES

Elsevier Services include an electronic preprint service in which articles accepted by the *Nuclear Physics* journals but not yet printed are made immediately available to subscribers over the network, shaving five months off the time it usually takes a paper to appear in print.

INSTITUTE OF ELECTRICAL AND
ELECTRONICS ENGINEERS, INC.
455 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
(908) 562-3998

IEEE

IEEE has had an arrangement (for some time) with UMI to deliver electronic versions of its multitude of publications (books, journals, proceedings, etc.) on CD-ROM. IEEE is conducting negotiations with an institution of higher education to deliver its documents over the Internet to faculty, staff, and

students. A number of user interfaces will be applied to the collection, and user-behavior will be studied to provide IEEE information about the most beneficial use of its publications in a research setting. Promulgation of documents among faculty and staff of engineering schools in this way will also help increase student membership and retention in the Society which sponsors and finances these publications.

THE FRENCH NATIONAL INSTITUTE OF
SCIENTIFIC AND TECHNICAL
INFORMATION, INIST
2, allée du Parc
de Brabois - 54514
Vandœuvre-lès-Nancy
Cedex France
+ (33 83) 50 46 40
+ (33 83) 50 46 50 (FAX)

INIST

While not—at the time of this writing—actually using the Internet as a delivery vehicle, INIST is clearly considering ways of doing so. INIST is a large (27,000 periodicals; 600,000 requests in 1992) traditional document delivery service as well as an abstracts database publisher ("Pascal" for STM has 10 million items). INIST is interested in pilot projects with publishers who would allow systematic scanning and electronic storing of the articles. However, even the scanned articles (so far, 1,700 periodicals are scanned via an arrangement with the French Reproduction Rights Organization) are almost always delivered by mail; only 4% is faxed, and nothing is yet electronically distributed. INIST claims to want electronic storage permission mainly to speed up this traditional fulfillment process. They are exploring electronic transmission, but this is only in embryonic stages.

INSTITUTE FOR SCIENTIFIC
INFORMATION
A Thomson Company
3501 Market Street
Philadelphia, PA 19104
(215) 386-0100

OCLC
Online Computer Library Catalog
6565 Frantz Road
Dublin, OH 43017
(614) 764-6000

ISI

At the time of this writing, ISI is not actively involved with the Internet. Reportedly, considerable attention has been given the possibility, however. ISI is a large secondary information publisher (*Current Contents*, *Citation Index*, et al.) and traditional document supplier ISI is moving towards scanning and OCR (optical character recognition) of the paper copies of the journals they cover, instead of just re-keying the abstracts, as ISI traditionally do. This would allow ISI the opportunity of creating electronic document delivery services. ISI is seeking agreements with publishers on pilot projects, which might eventually include using the electronic text as supplied by the publisher.

Mr Eric Maki
KLUWER ACADEMIC PUBLISHING
101 Phillip Drive
Assinippi Park
Norwell, MA 02061
(617) 871-6300

KLUWER ACADEMIC

Kluwer offers a complete Tables of Contents of its electrical engineering and computer science journals going back to issue 1:1, and including those of forthcoming issues. The FTP service is very easy to use.

OCLC

OCLC, a provider of library systems and services, including databases, are developing gateways to various document delivery services.

THE UnCover COMPANY
3801 E. Florida, Suite 200
Denver, CO 80210
(800) 787-7979
(303) 758-5946 (FAX)

UnCover

UnCover, a joint venture of CARL Systems and B. H. Blackwell, Ltd., is a current awareness and article delivery database covering around 15,000 periodicals. Journal Tables of Contents are loaded onto the Internet and orders are taken over the Internet. Articles are faxed or mailed. Libraries are the primary market. End-users may order articles directly, either with a credit card or with a library debit account. The database consists of tables of contents, available simultaneously with publication. Fulfillment is by FAX, either within an hour (in the event an article has been requested prior to publication and the publisher has given permission to "scan and store"), or within 24 hours, if it is a new request, or if electronic storing was not permitted by the publisher. UnCover can be reached via the Internet; no prior subscription is necessary. There is currently a service fee of \$8.50 plus a copyright fee, set by each publisher for every article delivered via UnCover.

B. Electronic Journals

There are an increasing number of emerging electronic journals on the Internet, many of which are society- or research-community-driven. Some are joint ventures with traditional paper publishers. The financing of many of these journals is somewhat in question. It is easier to launch an innovative electronic journal by faculty, for example, than it is to sustain and finance the effort over time. Those which have been undertaken in cooperation with established publishing houses might be expected to survive over time—but only if the financial and usage experience is supportive of ongoing efforts.

The following electronic journals seem worthy of note. There are numerous other examples that could have been included in such a compilation. There has been no effort to judge these examples as particularly meritorious or exceptionally innovative. They are included as examples of the kinds of efforts that are proliferating on the Internet and are journals that the professional and scholarly publishing community ought to watch with interest.

JOURNAL OF ARTIFICIAL INTELLIGENCE RESEARCH

mailto://jair@cs.cmu.edu Subject "autorespond"
Message "help", or
mailto://jair-ed@ptolemy.arc.nasa.gov

JAIR

The AI Access Foundation has started accepting submissions to *The Journal of Artificial Intelligence Research* [JAIR] in all areas of AI, including automated reasoning, cognitive modeling, knowledge representation, learning, natural language, perception, and robotics. JAIR is a refereed publication covering all areas of AI. In addition to being published electronically, each complete volume of JAIR will be published by Morgan Kaufmann.

From an e-publishing point of view, JAIR is notable because

- The journal is published both electronically and by a traditional publisher.
(Note: The publisher will be publishing each complete volume, not individual issues.); and

- The electronic version is free over the Internet (Articles are published in PostScript.).

Although JAIR has been publishing for only a few months, initial reaction suggests that the electronic version of the journal will be successful, if the number (and quality) of the submissions that have been received is any indication. Critical to the successful reception is that JAIR has enjoyed a readership consisting of computer scientists who generally have easy access to the Internet and are used to passing around PostScript articles over the it. As the network expands, and E-Publishing tools improve, journals like JAIR will presumably be possible in many scholarly communities and disciplines.

JAIR's success in attracting good submissions has also been aided by having a prestigious international editorial board, and by making sure its turnaround time (reviewing, etc.) is faster than any other journal in the field. The electronic form of JAIR also presents some added value over traditional journals. For instance, JAIR allows source code, experimental data, and demonstration programs to be electronically "published" in on-line appendices to articles.

The journal is run by a non-profit corporation, which has granted the publisher rights to publish the hardcopy volume. JAIR has been aided by several universities and research labs that have allowed JAIR to use their facilities as electronic archive sites. The publisher has also been very helpful.

JAIR is available over the Internet via Gopher, a USENET newsgroup, FTP, automated E-Mail, etc.

Its editor makes some interesting observations of interest to the PSP community:

"Basically, this journal can be distributed for free because almost everything is done by volunteers (the editorial staff), as part of their normal jobs as academics and researchers."

"Presumably, the publisher's revenue will come primarily from sales to libraries and individuals who don't want to rely completely on an electronic archive."

"Since we will basically give them [Morgan Kaufmann] a finished product to publish, they have very little in the way of costs. (We are not sure what the role of the publisher will be in 5 years, but we will be happy to work with them to see what added value they can provide.)"

OCLC
6565 Frantz Road
Dublin, OH 43017
(614) 764-6000

C. Publisher/University Pilots

The Online Journal of Current Clinical Trials

The Online Journal of Current Clinical Trials [OJCCT] was jointly developed by the American Association for the Advancement of Science [AAAS] and OCLC, with the latter providing the software. (Cf.: *Intro:F* to show recent approbation of OJCCT by the major bibliographic and citation vendor. This acknowledgment of its standards and values will go a long way toward permitting electronic publications to achieve the potential credibility of their paper counterparts.)

The first peer reviewed medical journal (started July 1992) delivers the text of articles within 24 hours after acceptance for publication to subscribers (\$110/year), either via modem or via the Internet. The journal has no paper counterpart. It does not contain halftones. It offers full text keyword searching, hypertext links to *Medline* abstracts, automatic availability of back issues, and letters, revisions, etc. linked to the original articles.

Despite these value-adding features and an attractive and sophisticated graphical user interface, developed by OCLC, OJCCT at the beginning was not attracting enough high quality papers from respected scientists. Scientists were reluctant to publish their work in an unproven medium and questioned especially, the functionality of citations. However, submissions are increasing now that OJCCT authors are given the option to have a short version of their paper considered for simultaneous print publication in the prestigious British journal *The Lancet* (separate peer review is required).

OCLC is planning to add additional electronic peer-reviewed journals using the same software as OJCCT. The *Online Journal of Knowledge Synthesis for Nursing* and an online version of *Electronics Letters*, the bi-weekly journal of the IEE have been announced. Skepticism exists about the feasibility of the latter, mainly because of complex technical aspects of the volume. It is just these kinds of challenges, however, that experimental prototypes should be addressing.

RED SAGE
Library and Center for Knowledge Management
University of California at San Francisco
San Francisco, CA 94143-0840
(415) 476-8293

Red Sage

The name of the three-year project jointly conducted by Springer-Verlag, AT&T Bell Labs, and the University of California at San Francisco (UCSF) is *Red Sage*, named after the landmark Washington, D.C. restaurant at which the final negotiations for the collaboration among these (seemingly unlikely) partners took place.

Red Sage is designed to explore some of the technical and behavioral issues associated with electronic delivery of primary journal information to scientists at their desks.

Springer-Verlag is providing twenty-four journals in electronic form in two subject areas (radiology and molecular biology). The scientists at the University of California will access the Springer journals with *RightPages™*, an innovative software program from Bell Labs that features electronic document browsing, text searching, and a system to alert users about new journal articles. The experiment which is now being extended to other publishers will run for three years.

1. How the System Works

After logging into the *RightPages™* system, the scientists at the University of California see an array of journal covers on their screens, similar to the display of periodicals in a library. To see which numbers of a particular journal issue are available on the system, the user need only "click" with the mouse on a displayed cover. After selecting an issue (again using the mouse) the table of contents is displayed. Articles in the issue are viewed by simply pointing to their entries in the table of contents. Journal issues can also be browsed page by page, and

articles can be printed on demand on local laser printers.

To help keep track of current publications, the RightPages™ system also alerts people to newly published articles in subject areas in which they have registered an interest.

2. Next Steps for Red Sage

Holding *Red Sage* together are the distinct informational needs about the functionality of the electronic environment each contributes to the collaboration. The primary objective of Springer-Verlag's participation in *Red Sage* is to explore some of the technical and behavioral issues associated with electronic delivery of primary journal information to scientists at their desks. AT&T wishes to learn about the applicability of their software to a distributed network environment and the problems encountered by increased and transient traffic, different user interfaces, and varied input standards. UCSF is exploring appropriate and satisfying mechanisms of fulfilling the information needs of its research community.

By June 1994 *Red Sage* expects to extend its reach across the Internet to other university campuses. Other publishers are beginning to understand that *Red Sage* could be important in the development of standards for electronic delivery and have asked if they can join the collaboration. At the time of this writing John Wiley & Sons, *The New England Journal of Medicine*, and the American Medical Association are joining the *Red Sage* project.

While the collaborators believe the project will go a long way towards defining the requirements for electronic delivery and usage, *Red Sage* is classed as an experiment, since no additional revenues will be generated during the three years of the project, and, indeed, all three organizations: Bell Labs, Springer-Verlag and the University of California at San Francisco currently participate financially in supporting the project.

ELSEVIER SCIENCE PUBLISHERS
655 Avenue of the Americas
New York, NY 10010
(212) 989-5800

TULIP

TULIP (The University Licensing Program) is a cooperative research venture between Elsevier Science Publishers and several major universities in the U.S. to provide electronic version of research journals and to test systems for networked delivery and use of electronic information.

The major goals of the project are

1. To determine the technical feasibility of networked delivery of electronic information. The participating universities will incorporate this material into local operational and prototype systems. A variety of delivery alternatives, search and retrieval systems and print-on-demand options will be used at the participating institutions. One of the goals is to compare these methods.
2. To understand, through the implementation of prototypes, alternative costing, pricing, subscription, and market models that may be viable in electronic distribution scenarios, and to compare these with existing print-based models.
3. To study reader usage patterns under the different distributions scenarios that are developed by the participating universities. Specifically, the project will examine:
 - ...the extent to which electronic access affects usage patterns;
 - ...the infrastructure required to support electronic access;
 - ...the printing needs and infrastructure needed to support them;
 - ...the differences in use between print and electronic format of material;
 - ...the costs of making material accessible electronically; and
 - ...how electronically published material should be best integrated with other resources such as online catalogs and journal abstracting databases.

Elsevier selected materials science as the discipline for this project and is providing participants

with bitmapped images of approximately 40 Elsevier and Pergamon journals in this field. Materials science was selected since the researchers in this area tend to be comfortable with computer applications, they have a higher than average installed base of workstations, and the publisher has a large body of frequently cited material in this area. The material starts with 1992 data. The project is expected to run until 1995, and represents approximately 125,000 journal pages a year.

D. Preprints, Table of Contents and Abstracts Databases

It is clearly difficult to separate the categories of these novel publishing projects. Readers will have to understand that the distinguishing characteristics of one form over another is less clear than these same characteristics in print.

Furthermore, the involvement of the readership in forming the future evolution of these electronic publications is yet to be fully understood. The network is bi-directional, which permits a closer interaction between publisher and consumer. It is likely that the reader/user community will have considerable and immediate impact on the evolution of electronic forms and enhancements. This constitutes a shift in *control* over publications which will involve learning adjustments for all the constituents in the process. These adjustments will have sociological, practical, and economic impact that are not entirely understood.

Rick Luce
Library Director
LOS ALAMOS NATIONAL LABORATORY
Research Library, MS P362
Los Alamos, New Mexico 87545
(505) 667-4448
<rick.luce@lanl.gov>

Physics Preprints Database

In 1991 Dr. Paul Ginsparg started an electronic bulletin board on which physicists could "post" articles destined for publication. Physicists can send preprints (as yet unpublished papers) to the bul-

letin board, and they can scan a list of preprints that are already on the system, read abstracts that could be of interest, and request entire articles to be sent to them electronically via the Internet (either by E-Mail, or via file transfer protocol [FTP]). Over 8,000 researchers now subscribe to one or more of the databases, each of which receives up to 600 articles per month.

The system is completely automated and unrefereed. Dr. Ginsparg stresses that inclusion in the database does not imply scientific acceptance of the contents or "validation." However, the databases are already the first port-of-call for physicists around the world for staying up-to-date.

Dr. Ginsparg does not deal with copyright problems; he claims the physics community has a long history of communicating by exchanging preprints. However, publishers may start to object, as it would seem that copyright is violated when articles are available on the Net at the same time as the transfer of copyright from the author to the publisher has taken place. So far, this issue is unresolved.

Note: a database of abstracts of math preprints is maintained at Yale University. The full text of the articles is available via FTP from computers at the various universities from which the preprints originated. The database is rumored not to have caught on in any major way, presumably because it is cumbersome to have to go to many different sources.

UNIVERSITY MICROFILMS
a Bell & Howell company
300 N. Zeeb Road
Ann Arbor, MI 48106
(313) 761-4700

UMI

UMI is beginning to explore using the Internet for delivery channels. Consideration is being given the possibility. As a major international vendor of information, such an eventuality could have significant impact on stimulating Internet usage and growth. University Microfilms International is a large secondary information publisher, mainly in the general, business, social sciences, physics and engineering areas. UMI pioneered the large CD ROM databases; it features abstract and index databases, full-text (ASCII) databases and full-image databases. With the last, a user can go directly from abstract searching to getting the image of the journal pages on a laser printer. UMI is now

offering multi-access systems, which allow libraries to use the image databases on their own networks or automation systems. Copyright payment is assured in UMI's system.

Director of Journal Information Systems
THE AMERICAN PHYSICAL SOCIETY
500 Sunnyside Boulevard
Woodbury NY 11797
Voice: (516) 576-2365
FAX: (516) 349-7817
E-Mail: rakelly@aps.org

American Physical Society Electronic Publishing Experiments

(From *APS Publishing News*, Issue 1 (October 1993) [reprinted with permission])

At the current time, the APS has three electronic publishing projects "in the works;" they are:

1. APS/Los Alamos National Laboratory Collaboration

An electronic version of past volumes of the *Physical Review* is being undertaken via a Wide Area Information Server (WAIS) in collaboration with the Center for Nonlinear Studies (CNS) at Los Alamos National Laboratory. The intent is to scan past issues for full-page displays to employ existing electronic files of past *Physical Review* volumes (from approximately 1987 to the present), and to make these files available in a form compatible with existing electronic search software (e.g., WAIS). Processes for incorporating equations and figures will be reviewed and an effective method for embedding them into the text will be developed. Beyond this six-year accumulation of *Physical Review* material, procedures will be developed for optically scanning the entire archive of *Physical Review* from its first issue, published in 1893. This is intended to be a two-year experiment when APS will gain experience in managing online journals and develop appropriate mechanisms for creating easy user access, charging algorithms, cross-referencing, etc. Gary Doolen (CNL), and Timothy Thomas (Computer Research) are co-principal investigators at Los Alamos.

2. APS/NRL Library Collaboration

An experiment is being initiated based on collaborative efforts between the APS and the Naval Research Laboratory (NRL) Library to scan past issues of *Physical Review Letters* and *Physical Review E* and storing them on optical disks. The technology for this project has already been developed and successfully implemented at the NRL Library. Users having access to the resulting database will be limited to members of the NRL research community via the Library's InfoNet. Access will be monitored in order to learn details concerning usage patterns of the APS journals. The experience gained will be used in future APS online initiatives. Laurie E. Stackpole, Chief Librarian, is the director of this project.

3. On-line Version of *Physical Review Letters*

A Request for Proposal (RFP) will be sent to suitable vendors requesting bids for the creation of an online version of *Physical Review Letters*. The APS is seeking a vendor who will work with the Society to deliver PRL online with a full range of enhancements, so that after several years the electronic product would become self-supporting. The Society is requiring that only current issues be made available, starting with the first issue of Volume 74 (July 1, 1994). The Society will consider auxiliary proposals in which sequentially produced volumes (beginning with Volume 73 and working backwards) would be made available electronically as well.

In this pilot program, the electronic product is to be produced concomitantly with the printed version, which will continue to be published unchanged. The electronic version will be portable and accessible to users through a variety of the most commonly used operating systems. Secondary products, such as CD-ROM versions, are to be available at a modest incremental cost. A preliminary version of the RFP has been sent to a select group of science librarians for comment and reaction before preparing the final RFP.

NUMERISCHE MATHEMATIK

Springer-Verlag
 Heidelberger Platz 3
 14197 Berlin
 Germany
 + 030-8207-1

Numerische Mathematik

The math journal *Numerische Mathematik* is offered in an electronic mode, in which the table of contents is provided before the print edition is published. The journal will be offered in a full text version beginning in February 1994. This data will be available as TeX or LaTeX files. There is a charge for the abstracts and full text which will be described in response to a "help" message.

SPRINGER-VERLAG NEW YORK, INC.

175 Fifth Avenue
 New York, NY 10010
 (212) 460-1501

Springer-Verlag

Springer-Verlag is offering an electronic look at the contents of selected Springer journals three to six weeks before the printed version appears. Specifically, the tables of contents, article titles, and abstracts will be available over the Internet. Scientists can select from the body of data, e.g. they may choose to receive information about specific journals.

An electronic "help" message to Springer, will initiate the process.

TELOS

Springer-Verlag Publishers
 3600 Pruneridge - Suite 200
 Santa Clara, CA 95051
 (408) 249-9314

TELOS - The Electronic Library of Science

A Springer-Verlag New York publishing imprint dedicated to merging the traditional print media

with the new electronic media. All of its publications are computationally oriented, and each TELOS publication is accompanied by, or associated with, an electronic component. This includes book/diskette, book/CD-ROM combinations, but also stand-alone CD-ROM products, and information delivered online as network applications. This conscious effort to be bi-modal amortizes the risk of network publishing, while gathering information about user behavior and community needs. It is also an effective way to build credibility in the community of authors who inhabit and utilize the network as their primary form of informal communication.

GLOSSARY

We are indebted to Hans J. Rocke <fzhrocke@othello.ucdavis.edu> for selecting the terms for the following glossary, and to Janice Kuta and Jim Connolly <JEK Associates, NY> for contributing many specifically publishing-related definitions.

Absolute Address
(machine address)

An address which indicates, in machine language code, the exact storage location where data or machine instructions are to be found.

Access

To find or store information in memory or on a peripheral device such as magnetic tape or disk. To communicate in some way with a device.

Access Method

Any of the data management techniques available to the user for transferring data from memory to an input/output device, or vice versa.

Access Time

The time interval between the instant at which data is requested to be retrieved or stored and the instant at which the operation is completed.

Acrobat

A set of software developed by Adobe Systems (PostScript) and used to produce files for distribution on networks. Files can include text, graphics, photos, and color, and can be exchanged irrespective of the application that created them. Acrobat creates new files in a format called PDF (*Portable Document Format*).

Address

A location on a disk or in memory within which a specific piece of information may be stored, or the number assigned to that location. *See also* Absolute Address, Indirect Address

ADEPT

Association for the Development of Electronic Publishing Technique.

ADMD

Administrative Management Domain.

Adonis	A collaborative effort by major publishers of scholarly material to provide bit-mapped electronic images of serials publications. Publications of this nature are restrictive and highly protected, intended primarily as a delivery mechanism to research libraries for internal use.	ASCII	<i>American Standard Code For Information Interchange</i> A character encoding standard which uses 7 of the 8 bits of a byte to define the codes for 128 characters. e.g., in ASCII, the number 7 is treated as a character and is encoded as 00010111. Because a byte can have a total of 256 possible values, there are an additional 128 possible characters that can be encoded into a byte, but there is no formal ASCII standard for those additional characters. Most IBM compatible PCs do use an IBM "extended" character set that includes international characters, Greek letters and mathematical symbols. <i>See also</i> EBCDIC
Anonymous Ftp	More than 1,300 systems make public files available using the login name "anonymous." Such files can be searched or transferred to your own system. The "anonymous FTP list" allows users to locate the files accessible through Internet.	ASR	<i>Automatic Send And Receive</i> , having the ability to receive data and produce it on a printer or to send it through a keyboard.
Apollo Bundles	Combination of HP Apollo 9000 series 700 workstations and X-terminal products; primarily for computer-aided design	Assembler	A program or set of programs that translate symbolic language programs to machine language programs, thus reducing much human effort in programming.
ARCHIE	One of the earliest on-line file-finding utilities, Archie maintains a database of the names of all files retained at public archive sites and is maintained monthly. Server computers are located throughout the world and if your Internet usage is restricted, you can access via E-Mail. Archie tells you where a file resides but doesn't allow you to search the file itself.	Asynchronous	Technically, having a variable time interval between characters, bits or events. Communication in this modes does not require that the recipient be present to receive the message. i.e., E-Mail, answering machines.
Archive	1. The storage of files (often in archived and compressed form) for future use, or the storage area which holds those files. 2. To create an archive, to move a file into an archive or to bundle multiple files together into a larger archive or library file.	Automation	The implementation of several processes by automatic means.
Argument	An independent variable which defines conditions of the preceding instruction or identifies conditions under which the instruction should take place	Backbone	A single-protocol connection among different systems, each one having a gateway to the common backbone protocol. These high-speed connections accept and distribute traffic to and from other networks.
ARPA	<i>Advanced Research Projects Agency</i> of the United States Department Of Defense.	Bandwidth	The difference (in hertz) between the upper and lower limits of wave frequencies transmitted over a communications channel.
ARPANET	This project began within the U.S. Dept. of Defense as an experiment in networking computers for research purposes. It was later split into two parts, Milnet (Military Network) and DARPA Internet (<i>Defense Advanced Research Projects Agency</i>) and this subsequently became known just as the Internet. A major development from ARPANET was TCP/IP, a set of programs that facilitates communication between computers.	Batch Processing	The processing of groups of jobs one after another without the need of or opportunity for operator interaction. Batch processing often ties up the computer or an area of it so that no other work may be done at the same time in that area. In some systems, the computer may interrupt an operating batch program to process another with higher priority.
Artificial Intelligence	Computer programs that perform functions, often by imitation, which are normally associated with human reasoning and knowledge.	Baud Rate	A measure of how fast data is transmitted between computers. Generally, a baud will equate to the transmission of a bit of data per second. Most common at present is 9600 baud and 14400 baud is available today. Baud rate can also be expressed in kilobits. i.e., 14.4 kilobits per second.

Binary	Number system with base 2 (two) using the digits 0 and 1. In base 2 the value 7 is expressed as 111 (or 3 bits), 1×2 to the 0 power, 1×2 to the first power and 1×2 to the second power. Software program files are written in binary format and cannot be displayed as characters. Software capable of understanding the particular binary format method of encoding information must be used to interpret the information in a binary formatted file. Binary formats are often used to store more information in less space than is possible in a character format file. They can also be searched and analyzed more quickly by appropriate software but are not normally portable, however.	CCA	1. <i>Common Cryptographic Architecture</i> ; IBM encryption software for MVS and DOS application. 2. <i>Compatible Communications Architecture</i> ; network equipment technology protocol for transmitting asynchronous data over X.25 networks. 3. <i>Communications Control Architecture</i> ; U.S. Navy network that includes an ISDN backbone called BITS (Base Info. Transfer System).
BIT	Contraction of <i>binary</i> and <i>digit</i> . A bit is the smallest unit of information that a computer can work with. Each bit is either a one or a zero. Often computers work with chunks of bits rather than one bit at a time; the smallest chunk of bits a computer usually works with consists of 8 bits or a byte.	CCITT	<i>Consultative Committee Of International Telephone And Telegraph</i> , Part Of The International Telecommunications Union, A UN Treaty Organization, Setting International Standards For Worldwide Telecommunication, e.g., X.25.
Bitmap	A digital representation of an image in which all of the dots or pixels making up the image correspond to specifically assigned bits in memory.	CCS	<i>Common communication support</i> ; SAA specifications for communications, which includes data streams (DCA, 3270) application services (DIA, DDM), session services (LU 6.2) and data links (X.25, token ring).
BITNET	<i>Because It's Time Network</i> . Developed in the early 80s, this network links research centers and academic institutions in over thirty countries. It is not part of Internet, but connections enable information, data and E-Mail to be transferred between Bitnet and Internet.	CD-ROM	<i>Compact disk read only memory</i>
Block	Physical data consisting of a fixed number of characters or records, and moved as a unit during transmission.	Chains	In Hypertext, linking randomly located material by means of address information included within the stored item, which cites the location of the succeeding and/or preceding item in the sequence.
Boolean	A system of logic devised by George Boole using a series of symbolic terms such as "and", "or", and "not" to express the relationship of data elements to one another.	Channel	Any communications pathway between two computers or between a terminal and a computer. It may refer to the physical medium, such as coaxial cable, or to a specific carrier frequency (subchannel) within a larger channel or wireless medium.
bpi	<i>Bytes Per Inch</i> . A measure of the storage density used on magnetic tape. A 2400 foot tape at 6250 bpi could theoretically store about 180 megabytes (million bytes) of data. In practice, however, one seldom find tapes with more than about 100-150 megabytes of data.	Character Encoding	A method of encoding characters including alphabetic characters (a-z, upper-case and lowercase), numbers 0-9, punctuation and other marks (e.g. Comma, period, space, &, *), and various "control characters" (e.g., Tab, carriage return, linefeed) using binary numbers. For a computer, for instance, to print a capital A or a number 7 on the computer screen, we must have a way of telling the computer that a particular group of bits represents an A or a 7. There are standards, commonly called "character sets," that establish that a particular byte stands for an A and a different byte stands for a 7. The two most common standards for representing characters in bytes are ASCII and EBCDIC.
Bulk Transfer	The transfer of a batch of data in a continuous burst by direct link between computers, or by magnetic tape transfer.		
Byte	1. Eight bits. A byte is simply a chunk of 8 ones and zeros. For example: 01000001 is a byte. A byte is equal to one column in a file written in character format. 2. Monthly periodical published by McGraw-Hill; ISSN 0360-5280		

Character Format	Any file format in which information is encoded as characters using only a standard character encoding scheme. A file written in "character format" contains only those bytes that are prescribed in the encoding scheme as corresponding to the characters in the scheme (e.g., Alphabetic and numeric characters, punctuation marks, and spaces. A file written in the ASCII character format, for instance, would store the number "7" in eight bits (i.e., One byte): 00010111. A file written in EBCDIC would store the number "7" in eight bits as: 11110111. <i>Contrast with Binary Format.</i>	Computer Literacy	Understanding computers and related systems. It includes a working vocabulary of computer and information system components, the fundamental principles of computer processing and a perspective for how non-technical people interact with technical people. It does not deal with how the computer works (digital circuits), but does imply knowledge of how the computer does its work (input, process, output). It requires a conceptual understanding of systems analysis and design, application programming, systems programming and data center operation. To be computer literate in management, an individual must be able to define information requirements effectively and have an understanding of what the primary decision support tools, such as query language, report writers, spreadsheets and financial planning systems can accomplish. In the case of the PSP community, the knowledge to apply traditional publishing skills and services in a distributed network medium.
Client	The user of a network service; also used to describe a computer that relies upon another for some of all of its resources.		
Client/Server	Architecture IUN which the client is the requesting machine and the server is the supplying machine (LAN file server, mini, or mainframe). The client provides the user interface and performs some or most of the application processing. The server maintains the database and processes requests from the client to extract data. The server also controls the application's integrity and security. Applications can be designed, whether running within the same computer or in multiple computers, in which one program (the client) requests data from another program (the server). For example, in X-Window, the server is the software that manages the display screen and the client is the application that asks the server to display something.	Connect Time	The time during which an operator is in contact with a computer online. This is different from compute time, in which the operator is actually utilizing the computer's resources.
		CORE	A joint project by OCLC, Cornell University, Bellcore and the American Chemical Abstracts service. CORE is similar to red sage but relies more heavily on the SGML component to identify manuscript elements as individual related "objects". The data for CORE have been published in traditional ACS journals.
CNI	<i>Coalition for Networked Information</i> The Association for Research Libraries (ARL), CAUSE and EDUCOM formed this organization to consolidate information in such areas as intellectual property rights, licensing, modeling, standards and similar subjects in information technology, libraries and higher education.	CREN	<i>Corporation For Research And Educational Networking</i> . A merger of BITNET and CSNET.
		CUPID	<i>Consortium For University Printing And Information Distribution</i> ; serving the community of university publishers and information distributors (CUPID) project: Harvard University, California State University System, Cornell University, Virginia Polytechnic Institute And State University, University Of Michigan, Princeton University, Pennsylvania State University, Duke University, University Of California at Davis, and Xerox Corporation.
Column	In a data file, a single vertical column each being one byte in length. Fixed format data files are traditionally described as being arranged in lines and columns. In a fixed format file, column locations describe the locations of variables.		
Compatibility	The ability of one device to interconnect or share programs or data with another by means of having the same code, speed, and signal level.	Cyberspace	Term coined by William Gibson in his novel "Neuromancer", to refer to a futuristic computer network that people use by plugging their brains into it.
Compress	Reduce the size of a file considerably by removing redundant information. Compressed files are more economical to transmit via Internet or to store. In order to use the file, it must be 'exploded', i.e., Reconstituted.	DASD	<i>Direct Access Storage Device</i>

Data	1. A general term for any collection of information, facts, numbers, letters, or symbols that refer to or describe an object, idea, condition, situation, or other factors. 2. Name of an android in <i>Star Trek - The Next Generation</i> .	DOS	<i>Disk Operating System</i> ; any of a number of widely used operating systems, so-called because a primary function they provide is the control of auxiliary storage in the form of disks.
Data Network	A telecommunications network built specifically for data transmission, rather than voice transmission.	Download	To transmit data from one central computer to another device or to a remote terminal.
Database	A set of organized data stored in, or available to, a computer, that may be used by the computer or its operator to perform various tasks. The database is not the program; rather it is the information the program will work with.	Driver	A program that controls (drives) the operation of a device or interface. The driver program interprets the computer data, providing the commands and signals required by the device or interface. The driver may output directly to the device or interface, or may provide paper output.
Datagram	The basic unit of information passed across the Internet. It consists of a source and destination address along with data. Large messages are broken into a sequence of IP datagrams.	Duplex	Pertaining to a transmission system where data may be received and transmitted. Half duplex - can only transmit or receive. Full duplex - can transmit and receive simultaneously.
Decryption	Encoding a message to its original meaningful form by means of a key.	E-Mail	<i>See Electronic Mail</i>
Desktop Publishing	A popular generic term describing a range of microcomputer-based typesetting systems that output camera-ready pages on plain paper xerographic [or laser] printers.	EBCDIC	A character encoding scheme used by IBM mainframe computers and some other computers. Unlike ASCII, the EBCDIC standard specifies use of the entire 8 bits of each byte. Example: in EBCDIC the number seven is treated as a character and is encoded as: 11110111. (EBCDIC stands for Extended Binary Coded Decimal Interchange Code.)
Disk Farm	Direct access storage devices (DASD) with online disk drive storage and gigabyte capacity used by banks and large manufacturing firms to insure rapid access to and updating of customer and other information.	Electronic Mail	A message service using electronics and telecommunications to deliver hard or soft copy information. These may take the form of text-only messages or of images which include text in font form and graphic material.
Document Delivery	A business providing full-text reproduced copies of articles or parts of publications. Of considerable interest to the PSP community because of the potential to avoid paying royalties. Conversely, an opportunity to provide greater service to a constituency who may not subscribe or otherwise purchase documents in any other way.	Encryption	An algorithm designed to protect the interpretation and use of intellectual property in electronic files against unauthorized use.
Domain	A part of the naming hierarchy essential for the correct identification of systems and networks and consisting of a sequence of names or other words separated by periods. e.g., @prep.ai.mit.edu	End User	The person or organization who will directly use a particular set of information or a device.
Domain Name System (DNS)	A global naming system for use in UNIX networking for general-purpose, name-to-resource mapping. While in Internet the Network Information Center (NIC) manages the higher-echelon domain names, the bulk of the management is decentralized to the lower-echelon sub-domains. Each <i>Domain Name Server</i> in an Internet community is responsible for the specific piece of the global name hierarchy over which it has authority.	Error	Any discrepancy between the theoretically correct behavior or values in a computer and the actual behavior or values. Most computers have routines specifically designed to detect the presence of errors.
		Escape	(ESC) A control code that indicates that the next code or codes has a different meaning than it would usually have.
		Ethernet	A communications protocol developed by Xerox Corporation, widely used for local area networks.
		Even Parity	<i>See parity</i>

Fiber Optics	A technology that uses light as a digital information carrier. Fiber optic cables are a direct replacement for conventional coaxial cables and twisted-pair wires.	Hard-Wired	Pertaining to the direct wiring of a terminal to a computer system (or any device to any device), as opposed to devices which communicate via telephone lines or wireless media.
File	A physical unit of storage on a computer disk or tape.	Hierarchical File	A file which contains information that is organized in subordinate levels or with a relational structure.
File Management	The activity of a computer in keeping track of what is happening to files, maintaining files, and allowing files to be edited.	Host	A system or subsystem in a network which performs actual processing operations against a database, and with which other network devices communicate.
File Transfer Protocol - (FTP)	An Internet program that allows users to connect to other computers to transfer programs and files but only in complete form.	Host Computer	The computer and associated database which run as a separate entity, but can be accessed through the network.
Firmware	Software that is stored in read-only memory (ROM). Firmware functions are not programmable by the user.	Host Site	The location (station) that receives communications from the other points in the network, performs operations on them via a host computer, and sends communications back to other points.
Fixed Format	A file structure consisting of physical records of a constant size within which the precise location of each variable is based on the column location and width of the variable.	HPPC	<i>High Performance Computing and Communication.</i> This initiative, developed by the Clinton Administration, focuses on the computing, communications and software technology needed to meet the accelerated need for information. HPPC will play a major role in the development of a National Information Infrastructure (NII), as outlined in the paper "Technology for America's Growth, a New Direction to Build Economic Strength," published on February 22, 1993. (A copy of the HPPC Act is to be found on the Internet at: ftp://ftp.nis.nsf.net . Other documents related to the Clinton Administration's Technology Initiatives may be found on the CNI file server at cni.org .)
Free Format	A physical file structure that specifies the order of variables in a file and that they are delimited from each other by a special character or characters (usually a blank or other whitespace). Free format files may have variable physical record lengths; when they do, they are typically delimited by a new line character at the end of each line. Contrast with fixed format.	Hypercard	A computer program storing text or multi-media in modules which may be addressed by their "handles" or "buttons", either from an outline or as part of the screen display. The information is arranged in "stacks" rather than in continuous sequence.
FTP	see File Transfer Protocol		
Full-Text	Referring to a database that contains entire documents as opposed to citations or abstract representations.		
Gate-Keeping System	A method of facilitating, as well as controlling the process of gathering, organizing, filtering, distributing and exchanging information in various formats between sources and users.		
Gateway	An electronic communications node which connects the individual user with specific mainframes or networks.	HYTELNET	A menu-based accessing system for MS-DOS users and one of the most thorough directories of telnet-accessible sites.
GOPHER	A software tool that allows users to retrieve information available from thousands of Gopher servers, anonymous FTP connections, publicly-accessible telnet connections, World-Wide-Web servers, Novell networks, WAIS servers and many more sources throughout the world. It is a menu-driven system and each Gopher server has links to every other Gopher server, allowing the user to move "seamlessly" from one server to another without the user being aware.	IF	<i>Interactive Facsimiles</i> , merges voice with facsimiles.
		Impact Printer	A printer that uses a physical mold for the character image, and impresses the image on the paper by striking a carbon or cloth ribbon against the paper with the character mold. The character masters may be on a daisy wheel, a metal band, a ball, or some other carrier.
Handshake	A protocol wherein a transmitting device sends signal, then the receiving device sends a "ready" signal before transmission continues.	Indirect Address	An instruction that references an address specified in the content of another address.

Inkjet Printer	A printer that uses finely directed sprays of ink to produce the character image.	Laser Printer	A printer that uses a laser to xerographically generate the image to be reproduced. Laser printers are typically the most sophisticated printers, often capable of producing both character and graphic images via raster image processing technology, often rivaling typographic quality. They are the output device usually associated with desktop publishing.
Input	1. The process of entering data via a keyboard or terminal or other device. 2. The material that is entered via keyboard or terminal.		
Intelligence	The ability of a device to make computational and evaluative decisions under the control of a program.	ListServ	A feature of E-Mail that allows a single message to be "served" or distributed to many addresses simultaneously. The recipients' E-Mail address must appear on a distribution "list" in order to receive these reflected messages. ListServ also provides the facility for "computer conferencing" with other users, all having a common interest.
Interactive	1. Relating to the ability of a device or procedure that allows an operator to make decisions that influence the outcome of a procedure in process. 2. Pertaining to a device that allows an operator to input data or commands and which then responds in some way to the operator.		
Interface	The point and manner in which two separate systems or devices connect and interact with one another.	Local Area Network (LAN)	A collection of devices and communication channels that connects a group of computers and peripheral devices together so that they may communicate with each other. Typically, local area networks occupy a single building or office area.
INTERNET	A hierarchical collection of thousands of TCP/IP campus, state, regional, national and international networks into one single logical network, all using a common addressing scheme.	Log	1. A record of operations on a file, indicating actions such as file creation, modification, errors, and other data. 2. To sign on (log on) or off (log off) a computer system or area. Log on procedures may require operator passwords, or may be accomplished simply by designating the desired area.
IP	<i>INTERNET protocol</i>	Logical Record	All the data for a given unit of analysis. It is distinguished from a physical record because it may take several physical records to store all the data for a given unit of analysis.
IP Address	Just as each computer in a network has a name identifier, each also has an address, in which four numbers are separated by periods. i.e., 271.86.134.78. In most situations, the user does not need to know the IP Address since this is located automatically by the host computer.		
ISDN	<i>Integrated Services Digital Network</i>		
ISOC	<i>The Internet Society</i> . A membership organization which supports the international information network.	Macro	Describing a computer instruction that has a "name" rather than an explicit code, and that will be translated or expanded by the computer to a set of instructions.
Kernel	The level of an operating system or networking system that contains the system level commands or all of the functions hidden from the user. In a UNIX system, the kernel is a program that contains the device drivers, the memory management routines, the scheduler, and system calls. This program is always running while the system is operating.	Main Memory	The memory space that is located in the central processor and accessed directly by the computer, as opposed to memory located on peripheral devices such as discs or tapes.
		MAPI	<i>Message Application Programming Interface</i>
KNOWBOT	An information retrieval technique developed by Knowbot Information Service (KIS) that understands how to navigate networks to locate specific information.	MD	<i>Management Domain</i>
		Melvyi®	A centralized information system for all nine campuses of the University of California. It includes a library catalog database, a periodicals database, article citation databases, and other files and can be accessed via UC lines or the internet.
LAN	A <i>Local Area Network</i> , such as within a city. <i>See also</i> Local Area Network.		
Laser	<i>Light Amplification By Stimulated Emission Of Radiation</i>		

Memory	The internal storage capacity of a computer system. Memory is generally located on some magnetic device such as disk, drum or core. Data is stored in digitally encoded bytes, and manipulated as needed during calculation processes. The amount of memory a computer has directly affects its ability to perform complex functions.		
MHS	<i>Message Handling System</i>		
MILNET	<i>Military network</i>		
MIME	<i>Multipurpose Internet Mail Extensions.</i> A code to specify the format of Internet messages, identifying text, program, mail- message, image, audio, video, and/or multi-part (mixed media) files.	Mouse	A handheld input device, similar to a keyboard, but limited to such interactions with graphic images on the screen as pointing, clicking and dragging.
Mobidem	Mobile modem for wireless data communication	Multi-Tasking	Describing a computer's ability to carry out two or more functions simultaneously.
Mode	A particular condition or state under which a computer or other device may operate, such as an insert mode, a communications mode, or a binary mode. Operations or commands may take on different meanings in different modes.	Multiplex	To transmit simultaneously two or more messages over the same communications channel to different receivers.
Modem	<i>Modulator/demodulator</i> , connects computers at remote sites to the telephone system by converting data from the host computer into electronic signals which are transmitted via the network to the target computer where another modem converts them back into machine-readable data. Modems can send as well as receive data, enabling computers to "converse" with one another.	Network	Two or more computers linked together physically or via telecommunications for the purpose of electronically sharing resources such as files, programs, and peripheral devices.
MOSAIC	1. A free Macintosh client browser for World Wide Web servers, developed at the National Center for Supercomputing Applications in Champaign, IL. The program can access linked data on Internet servers via many protocols: Archie, Gopher, Wide Area Information Servers, Ftp (File Transfer Protocol), Telnet and Network News Transfer Protocol. The center in Champaign, a Federally-funded research site, is receiving over 600,000 electronic information requests each week from Mosaic users. 2. The Macintosh version of Mosaic requires System 7 and MacTCP 2.0.2. Mosaic is available via anonymous ftp from FTP.NCSA.UIUC.EDU in the directory /mac/mosaic. 3. Netware for Mac developed by Novell inc., allows users to share non-Postscript printers connected to the PC portion of a netware network. The software intercepts the Postscript print job and translates it into PCL (printer control language) before sending it to the netware server's print services.	Network Control Program (NCP)	A program within the software of a data processing system which controls the performance of a telecommunications network.
		Network File System (NFS)	A process for mounting magnetic disks on a network so that disks not physically attached to a computer appear as if they were physically attached.
		NIC	<i>Network Information Center.</i> An organization that provides management of and training in, a particular network.
		NII	<i>National Information Infrastructure.</i> A program developed by the Clinton Administration to build the nation's computing and communications infrastructure to meet the needs of the next century.
		NFS	<i>Network file system.</i>
		NNSC	<i>National Science Foundation Network Service Centre.</i> The Center provides an Info-Server for files accessed by anonymous FTP, as well as being the help and service center for NSFNET.
		Node	A branching or exchange point in a network.
		Noise	Unwanted signal or signals on an electrical circuit.
		Noise Immunity	The ability of a device to accept valid signals while rejecting invalid signals.
		Non-Switched Line	A communications link permanently installed between two points.

NREN	<i>National Research And Education Network.</i> A developing fiber-optic (improved capacity and transmission speed) network to link all education facilities, including grade schools and libraries, institutes of higher education, and government organizations. This network should be completed in the mid-90s, but to an extent exists now as NSFNET and is referred to as Interim NREN.	Optical Disk	A rigid, plastic disk, 4 3/4 " in diameter, with an embedded metallic underside on which data are recorded permanently by laser (ROM - read-only-memory). Also known as compact disk, CD or CD-ROM. Digital information is recorded on a master disk with a strong laser beam. Copies are made by "stamping" 4 3/4 inch disks. They are read by a weaker laser beam. It is a high-density storage medium, with a capacity of ca. 600 megabytes.
NSFNET	<i>The National Science Foundation Network.</i> Funded by the National Science Foundation and operated by the Merit Corporation, this national backbone network is used to interconnect regional (mid-level) networks such as WestNet to others.	Optical Scanner	A device that uses light to scan and convert text, graphics, or other visual images into digitized data that can be read by a computer.
Null Modem	The communications cable used in hardwiring.	OSI	<i>Open System Interconnection</i>
Numeric Database	A database primarily containing numbers.	Packet	An addressed data unit of convenient size for transmission through a network.
OCR	See Optical Character Recognition	Parallel Interface	A data transmission technique in which a group of binary digits (bits) are transferred simultaneously over multiple lines. Usually, eight bits that correspond to a character are transferred as a single operation.
Octet	The grouping of eight numbers in a pair and two triplets as used in the domain name system; e.g., 35.222.222.	Parameter	1. A designation for the format of type, as requested by command codes or system defaults. Line length, page depth, typeface, and leading are examples of parameters. Commands which take place as they are entered and do not maintain their effect (such as extra leading, cancel escapement, etc.) are not parameters. 2. (<i>program parameter</i>) a constant or variable which remains unchanged in a subroutine, and fully or partly specifies a process to be performed during the subroutine. 3. (<i>hardware parameter</i>) a parameter characteristic of a machine which establishes certain limitations or capabilities of the machine. For instance, a parameter of a typesetter might be its ability to process 160 lines per minute.
Odd Parity	See parity		
Offline	Pertaining to a device or function which is not electronically connected to the main device. Media transmission from an offline device must be by means of manually carried material (disk or tape) or via telephone lines.		
Online	Pertaining to devices that are electronically connected to the computer. Generally, an online device is treated as if it were an integral part of the computer system.		
Online Public Access Catalog (OPAC)	A database of bibliographic records to which the public has access, reflecting the material owned by a library or a consortium of libraries.		
OPAC	<i>Online Public Access Catalog</i>	Parity	The transmission of data so that all codes have either an even or an odd number of one-bits. Even parity means, that to the codes with an odd number of one-bits one bit has been added so the total is even, and odd parity means that all the even codes have a bit added to make them have an odd number of one-bits.
Operating System	The program or set of programs which control a computer's operations and monitor the functions of the other programs.	Parity Check	A method of verifying the accuracy of a transmitted bit pattern by examining the code to determine whether its value is odd or even.
Optical Character Recognition (OCR)	A method of converting graphic symbols (particularly alphanumeric) into electronic signals by means of a reading device that recognizes character shapes. Until recently, material prepared for optical character recognition had to be typed in a specific format, and with a specific type element.	Password	An identification number keyed by the operator and checked by the computer before the database may be accessed.

Pay-By-The-Drink	A method of charging the user every time a file is accessed, based on the number of connect time units elapsed and/or records inspected. This is in contrast to paying a subscription or license fee which permits repeated and unlimited access to the file(s) and records for the duration of the stipulated period of time.	Program	Series of instructions to direct the computer to perform specific tasks in a certain order.
PDF	<i>Portable Document Format. See also Acrobat.</i>	PROM	<i>Programmable Read-Only Memory.</i> A type of memory that can be programmed after the time of manufacture, but once programmed cannot normally be changed. A certain type of prom which is programmed by exposure to ultraviolet light may be reprogrammed with special equipment.
Peripheral	A device that is external to the system processor, but operates under the processor's control, such as a line printer or a communications signal.	Prompt	An on-screen processing technique that questions or "prompts" the user of a computer system for responses.
PERL	<i>Practical Extraction And Report Language.</i> A 'super-language' and 'kitchen sink language' combines the best features of C, SED, AWK, shell programming, database access and text manipulation. PERL operates via vectors and arrays, which permit it to carry out AWK-like functions, such as split, join, reverses and sort. It also supplies many string functions with operations similar to those of C's string functions. PERL has standard file-manipulation system calls which are accessed as built-in functions, as well as other facilities which ease manipulation of data base management style files. PERL's useful attributes allow users to 'massage' data, manipulate files and perform general small-task programming with ease.	Protocol	The proper procedure and sequence of events for data transmission of a particular input device, with regard to code structure, identification of the text stream, and code recognition.
		Query	A data message structured so as to elicit a response from a computer.
		Queue	A group of items arranged in a particular sequence, waiting to have some process performed.
		RAM	<i>Random Access Memory,</i> which see.
		Random Access	Access to data, information or files without observing any particular order.
		Random Access Memory (RAM)	A form of volatile memory which allows data (such as documents) to be stored randomly and retrieved directly by an address location. The system accesses the addressed material, with no need to read through intervening data. Information may be retrieved more speedily from random access memory than from serial media such as tape.
Platform	A manufacturer's operating system for functions that have not been standardized in the industry and require intermediary programs to interpret the commands across different systems.	Raw Data	Data which has not been processed, reorganized, or manipulated.
Port	A communications channel between a computer and another device, such as a terminal, modem, or printer.	Read Only Memory (ROM)	Memory that is programmed at the time of manufacture, it may be accessed only and cannot be erased. Being non-volatile, read-only memory holds its contents after the power is shut off. It can be used to contain an operating system, language translators, and other "permanent" software.
Postscript	Software made by Adobe Systems for desk top publishing, using a page description language which many laser printers understand.	Real Time	Pertaining to the performance of a computer in such a way that the operator receives responses quickly enough so that there is no effective delay in the operator's activity.
PRIMIS	McGraw Hill Inc.'s customized electronic publishing program for the college textbook market.	Record	In a database, a single related grouping of data items (such as a company name, address, and telephone number).
Printer	Peripheral device, connected to a computer to render images on paper in black/white or color. Pigments can be transferred from a ribbon pressed against the print surface by an array of wires (matrix or impact printer), heated to sputtering and directed through tiny holes (ink jet) or electrostatically attached to a metal drum and transferred to paper by the xerographic method (laser).	Red Sage	UC San Francisco/Springer Verlag/AT&T Bell Laboratory collaboration to publish an electronic journal in radiology and molecular biology using Bell's RightPages software.
Printout	Display of file contents in text or graphic form on paper.		

Relational Database	Data are stored in tables (similar to a spreadsheet) and the user can manipulate them in order to prepare specialized reports.	Site License	Authorization (usually subject to a fee) to have multiple copies for simultaneous use within a specified organization or location.
Remote	Pertaining to communication with a device located at some distance from the central device, but connected in some way, with cables, wires, or telephone hookup.	Siteprint	A service offered by dialog to transmit the data of database search results to a local printer (LaserJet) for printing on site.
Remote Login	An interactive tool allowing a user to connect to another computer, either with cable, wires or telephone hookup. In Internet, this is done with the telnet command.	SLIP	<i>Serial Line Internet Protocol</i>
Rich Text Format	(RTF) A text exchange standard proposed by Apple to provide an application independent file format including fonts, tab positions, rulers, line breaks, hyphenation, paragraph spacing, embedded objects, such as pictures and sounds, and special text styles.	Soft Return	A line-feed/carriage return inserted and removed automatically by text processing software to create variable word wrap.
ROM	<i>Read-only-memory.</i>	Software	A term coined to contrast computer programs with the hardware of a computer system. Software is a stored set of instructions that governs the operation of a computer system and makes the hardware run.
Router	A dedicated computer that sends packets from one place to another, paying attention to the current state of the network.	Stand-Alone	Referring to a device which is capable of performing the functions for which it is designed without the aid of or connection to another, smarter device. The stand-alone device may receive media from another device for processing.
RS-232 Port	A standard plug with 25 pins, used to connect computers and i/o devices.	Standard Generalized Markup Language	(SGML) The ISO standard language that enables data to move between media by describing documents by their structural elements rather than their visual format and thus permitting further analysis or reuse by various application programs.
RTF	<i>Rich Text Format</i>	Star	A network configuration in which a central controller communicates directly with each device or station. A diagram of this configuration looks like a star.
Screen	The display surface of a video terminal or cathode ray tube.	Static Memory	Memory that does not require refreshing.
Scroll	A function available on most video terminals where the display image seems to move up and down (or left and right) on the screen to provide incremental viewing of material earlier or later in the file.	Station	One of the input or output points of a communications system or of a multi-user computer.
Search	The electronic comparison of character strings entered by the operator with those contained in files or databases, arranged in boolean logic expressions.	Stop Bit	In asynchronous communication, a marker following each character.
Server	A network computer which shares its resources, such as files and printer, with other computers. e. g., Network file system (NFS)	String	A sequence of entities, such as characters or commands.
SFQL	<i>Structured Full-Text Query Language</i>	Structured Query Language	(SQL) [pronounced 'sequel'] a data access language designed to work with relational databases. First used on IBM's DB2, SQL became a de facto standard in the mid-1980's.
SGML	<i>Standard Generalized Markup Language</i>	Synchronous	In data transmission, refers to the ability of the sending and the receiving devices to run continuously at the same frequency with a regular or predictable time relationship..
Sign On	To connect with a remote computer by providing identification details and performing appropriate procedures.		
Signal	The electrical quantity that conveys data from one point to another.		
Simplex	A modem that either sends or receives information but cannot do both during one transmission.		
Simultaneous Transmission	The transmission of data in two directions at one, both sending and receiving, by the same device.		

System	A machine or devices using various hardware and software to accomplish certain tasks. Commonly composed of a central processor, with one or more input or output devices, and capable of making decisions about the material which is being processed. Various fields of computer technology tend to define "system" in slightly different ways.	Token Sharing Network	A communications network designed so that all computer devices (stations) on the network are connected to a common channel, called a bus. A group of bits (a token) is passed from station to station to give each station in turn the opportunity to transmit data.
TCP/IP	<i>Transmission Control Protocol/Internet Protocol</i>	Transmission Control Protocol	(TCP) a set of protocols, resulting from ARPA efforts, used by the Internet to support services such as remote login (TELNET), file transfer (FTP) and mail (SMTP). A set of programs that enable communication between similar or dissimilar computers on a network.
Tele-communication	The transmission of signals by telegraph, radio, satellite, or some other means that does not involve physical connection between the sending and receiving devices.	Transparent	Pertaining to a process that is thoroughly compatible with another process, so that a user is not necessarily aware that there is more than one process or function involved. A computer program that can be added to an existing program without retraining or re-educating its users is transparent to the original program.
Teleprocessing	Computer operations carried out via long-distance communications network.		
Teletext	A generic term for one-way information retrieval systems that broadcast digitally encoded text and graphics to remote users. Teletext users request pages of transmitted data by means of a keyboard.	Tree	1. (tree structure) an arrangement of data in a hierarchical form, with each group of data containing subgroups that present more detail. 2. A network configuration in which a central controller communicates to a number of other devices or stations, each of which on turn communicates to another group of devices or stations, in an hierarchy.
TELNET	The Internet standard protocol for remote terminal connection service. TELNET allows a user at one site to interact with a remote timesharing system at another site as if the user's terminal were connected directly to the remote computer.		
Terminal	A device (usually with a video display) on which an operator may communicate with or receive communication from a computer.	TULIP	<i>The University Licensing Program</i> a cooperative research project among U.S. Universities and Elsevier Science Publishers to distribute electronic journals. UC is planning to load bit-mapped images on a mass storage server and link them to relevant records in article databases on the melvyl system. A test mode is projected for the third quarter of 1993.
Terminal Emulation	A software capability in which a computer or terminal can be made to simulate the characteristics of another terminal for communications compatibility.		
Terminal Server	A small, specialized networked computer that connects many terminals to a lan through one network connection. Any user on the network can then connect to various network hosts.	Turnkey	Pertaining to any product that is delivered complete with any special hardware or software required, ready to run, without special onsite programming.
Text	A series of words or characters having some meaning to the reader, as opposed to command codes or instructions.	UNIX	A computer operating system developed by Bell Laboratories, written in the C programming language, and distinguished by its portability to different computers. It is widely used in graphics workstations.
Throughput	The net speed of a device or system, including input, output, and processing speeds together. The throughput speed may be slower than the inherent input or output speeds, since input may be slowed by simultaneous output, or vice versa, or input and output may have to take place serially.	User Friendly	A characterization of computer products that are easy to learn or to use.
Timesharing	The use of a computer for two or more purposes during the same time interval. The computer shares its attention among the devices by means of some monitoring program.	USENET	<i>User's Network.</i> A huge cooperative network with over 2.5 million participants worldwide and at least 10,000 hosts, the network supports informal discussion groups on topics of common interest. It is generally accessible through Internet and could be considered a computer conferencing network.

Value-Added	Enhancement of documents or data to increase their usefulness, such as gate-keeping and validation (i. e., by peer review), substantive and copy editing by publishers. Libraries add value by selection, cataloging, archiving and providing access. Electronic publishing adds value by providing full-text searching, audio, video, animation. Providing large data sets, reducing the time required for publication, and ease of access from workstations.	WWW	World Wide Web. Designed as an information retrieval mechanism, WWW is a hypertext interface to Internet resources. It has greater capabilities than Gopher in managing documents and presents information in a hypertext format rather than as a series of menus. Hypertext links may lead you down a search path that is fruitful or that may be a dead end.
VERONICA	Based on the ARCHIE model, VERONICA search its master index using keywords to access menu offerings on different systems in Internet, locates the requested data and makes the necessary connection.	WYSIWYG	What You See Is What You Get.
Video Display Terminal	(VDT) an operator station that includes a display screen as part of its hardware. VDTs at one time were called CRTs <i>cathode ray tubes</i> .	X.25	CCITT standard (1976) for the protocols and message formats that define the interface between a terminal and a packet switching network.
Virtual Library	Extends resources through access to bibliographic databases, full text files, images and other information in electronic format.	X.400	Global electronic messaging architecture
Voice Grade	Referring to the capability of a data transmission circuit to permit a transfer rate of up to 2400 bits per second (BAUD). A voice-grade circuit uses analog phone lines identical to standard telephone equipment, and encodes digital data into tones for transmission.	X-Windows	A device-independent, multitasking windowing and graphics system, commonly used in a UNIX environment.
Voyager Expanded Book	Electronic books produced by using voyager software for use on Macintosh PowerBook computers. They include enhancements such as sound, illustrations, cross-referencing and more. <i>Jurassic Park</i> is the best-known example; such editions may be produced and marketed by the voyager company or by traditional publishers under software licensing agreements.	Z39.50	Application-layer protocol standard developed by the American National Standards Institute in 1992 for computer-to-computer information retrieval. A number of companies and universities are using it to develop interoperable search-and-retrieval software products for accessing Internet. The Open Systems Interconnection (OSI) model defines Z39.50, but because OSI is not built into the UNIX operating system like TCP/IP, Z39.50 was adapted to run over TCP/IP for the Internet.
VT-100	A terminal developed by Digital Equipment Corporation (DEC) used as an access standard for text-only videotex systems.		
WAIS	Wide Area Information Server		
Wide Area Information Server (WAIS)	A sophisticated indexing tool using keywords to search hundreds of files and then index the relevant ones into a single .SRC file. New resources are announced in a special indexed database called "directory-of-servers."		
Wide Area Network (WAN)	A network spanning hundreds or thousands of miles, in contrast to a local area network.		

SELECTED BIBLIOGRAPHY

We are indebted to Jim Connolly of JEK Associates for the following essential bibliography.

Charles Bailey, Jr, University of Houston
<LIB3@UHUPVM1.bitnet> also made available
his now-definitive bibliography which is avail-
able as a separate handout.

Cronin, Mary. *Doing Business on the Internet*. New York, NY: Van Nostrand Reinhold, 1994.

Derry, Daniel. *The Internet Guide for New Users*. New York, NY: McGraw Hill, 1993.

Engst, Adam. *Internet Starter Kit for the Macintosh*. Indianapolis, ID: Hayden, 1993.

Estrada, Susan. *Connecting to the Internet: An O'Reilly Buyer's Guide*. Sebastopol, CA: O'Reilly & Associates, 1993.

Fraase, Michael. *The Mac Internet Tour Guide: Cruising the Internet the Easy Way*. Chapel Hill, NC: Ventana Press, 1993.

Gilster, Paul. *The Internet Navigator: The Essential Guide to Network Exploration for the Individual Dial-Up User*. New York, NY: John Wiley & Sons, 1993.

Hahn, Harley. *Internet: The Complete Reference*. New York, NY: McGraw Hill, 1993.

Hesslop, Brent. *The Instant Internet Guide: Hands on Global Networking*. New York, NY: McGraw Hill, 1994.

Kehoe, Brendan P. *Zen & the Art of the Internet*. New York, NY: Prentice Hall, 1992.

Krol, Ed. *The Whole Internet User's Guide & Catalog*. Sebastopol, CA: O'Reilly & Associates, 1992.

LaQuey, Tracey and Jeanne C. Tyer. *The Internet Companion: A Beginner's Guide to Global Networking*. Reading, MA: Addison-Wesley, 1992.

Lynch, Daniel and Marshall T. Rose. *Internet System Handbook*. Reading, MA: Addison-Wesley, 1993.

Malamud, Carl. *Exploring the Internet: A Technical Travelogue*. New York, NY: Prentice Hall, 1992.

Marine, April, Editor. *Internet: Getting Started*. Menlo Park, CA: SRI International, 1992.

Que Development Group Staff. *The Hitchiker's Guide to Internet*. Indianapolis, ID: Que, 1993.

DRAFT

The Internet Index

[Inspired by "Harper's Index"***]

Compiled by Win Treese (treese@crl.dec.com), 7/8/93; Revised: 1/7/94

Annual rate of growth for Gopher traffic: 997%
Annual rate of growth for WWW traffic: 341,634%
Average time between new networks connecting to the Internet: 10 minutes
Number of newspaper and magazine articles about the Internet during the first nine months of 1993: over 2300
Number of on-line coffeehouses in San Francisco: 18
Cost for four minutes of Internet time at those coffeehouses: \$0.25
Date of first Internet mail message sent by a US President: 2 March 1993
(Sent by Bill Clinton, President of the United States)
Date on which first Stephen King short story published via the Internet before print publication: 19 Sept 1993
Number of mail messages carried by IBM's Internet gateways in January, 1993: about 340,000
Advertised network numbers in October, 1993: 16,533
Advertised network numbers in October, 1992: 7,505
Date after which more than half the registered networks were commercial: August, 1991
Number of Internet hosts in Norway, per 1000 population: 5
Number of Internet hosts in United States, per 1000 population: 4
Number of Internet hosts in October, 1993: 2,056,000
Round-trip time from MIT to mcmtvax.mcmturdo.gov in McMurdo, Antarctica: 640 milliseconds
Number of hops: 18
Number of USENET articles posted in two weeks during December, 1993: 605,000
Number of megabytes posted: 1450
Number of users posting: 130,000
Number of sites represented: 42,000
Number of Silicon Valley real estate agencies advertising with Internet mail addresses: 1
Terabytes carried by the NSFnet backbone in February, 1993: 5
Number of countries reachable by electronic mail: 137 (approx.)
Number of countries not reachable by electronic mail: 99 (approx.)
Number of countries on the Internet: 60
Amount of time it takes for Supreme Court decisions to become available on the Internet: less than one day.
Date of first National Public Radio program broadcast simultaneously on the Internet: 21 May 1993
Percent of Boardwatch Top 100 BBS systems with Internet Connectivity: 21
Number of people on the Internet who know you're a dog: 0

***"Harper's Index" is a registered trademark of Harper's Magazine Foundation.
Copyright 1993,1994 by Win Treese. Reprinted by permission.
Send updates or interesting statistics to treese@crl.dec.com.

Published by The Electronic Information Committee
of The Association of American Publishers,
Professional and Scholarly Publishers Division

Robert Badger	<i>Springer-Verlag</i>
Nick Baker	<i>Elsevier Science Publishing</i>
Susan Barclay	<i>American Chemical Society</i>
Rick Bowes	<i>Macmillan New Media</i>
Laura Conley	<i>John Wiley & Sons</i>
Gary Goldstein	<i>Matthew Bender & Co.</i>
Cathy Guldner	<i>Oxford University Press</i>
Phyllis Hall	<i>IEEE</i>
Stephen L. Haynes (Chairperson)	<i>West Publishing Co.</i>
Ted Hill	<i>Simon & Schuster</i>
Isabella Hinds	<i>Copyright Clearance Center</i>
John Jenkins	<i>Bureau of National Affairs</i>
Timothy King	<i>John Wiley & Sons</i>
Charlotte Kuenan	<i>Bureau of National Affairs</i>
Taissa Kusma	<i>American Institute of Physics</i>
Barbara Polansky	<i>American Chemical Society</i>
Blake Ratcliffe	<i>University Microfilms Intl.</i>
Bernard Rous	<i>Association of Computing Machinery</i>
Carol Risher	<i>Assoc. of American Publishers</i>
Linda Scovill (PSP Division Chairperson)	<i>Scovill, Paterson</i>

AAP / PSP

Promises & Pitfalls
An AAP/PSP Briefing
Paper on
Internet
Publishing

An Adjunct Bibliography

A definitive bibliography presenting paper and electronic sources to help understand network based publishing.

Charles W. Bailey Jr., University Libraries, University of Houston
"Electronic Publishing on Networks: A Selective Bibliography of Recent Works."
The Public-Access Computer Systems Review 3, no. 2 (1992): 13-20.

AAP/PSP expresses its gratitude for the author's permission to reproduce this bibliography for use with the "AAP/PSP Briefing Paper on Internet Publishing."
February, 1994

+ Page 13 +

 Bailey, Charles W., Jr. "Electronic Publishing on Networks: A Selective Bibliography of Recent Works." The Public-Access Computer Systems Review 3, no. 2 (1992): 13-20.

1.0 Introduction

Despite a variety of problems, electronic publishing on BITNET, Internet, and other networks is experiencing vigorous growth as scholars experiment with FTP archives, list servers, WAIS servers, and other technological tools in order to reinvent scholarly publishing.

This bibliography presents selected sources, in both paper and electronic form, that are useful in understanding network-based electronic publishing. A limited number of sources that deal with broader electronic publishing topics, such as intellectual property rights, multimedia systems, standards, and virtual libraries, are also included; however, this bibliography does not provide an in-depth treatment of the large and diverse body of literature that deals with electronic publishing as a whole. In order to focus on recent developments, it does not cover sources published prior to 1989 (most sources are from 1990 to the present).

Hopefully, this bibliography will introduce interested readers to sources that will provide them with insight into the incredible intellectual ferment associated with network-based electronic publishing.

2.0 Bibliography

Alexander, Adrian W., and Julie S. Alexander. "Intellectual Property Rights and the 'Sacred Engine': Scholarly Publishing in the Electronic Age." Advances in Library Resource Sharing 1 (1990): 176-192.

Amiran, Eyal, Elaine Orr, and John Unsworth. "Refereed Electronic Journals and the Future of Scholarly Publishing." Advances in Library Automation and Networking 4 (1991): 25-53.

Amiran, Eyal, and John Unsworth. "Postmodern Culture: Publishing in the Electronic Medium." The Public-Access Computer Systems Review 2, no. 1 (1991): 67-76. (To retrieve this article, send an e-mail message that says "GET AMIRAN PRV2N1 F=MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)

Arms, William Y. "Scholarly Publishing on the National Networks." Scholarly Publishing 23 (April 1992): 158-169.

+ Page 14 +

Bailey, Charles W., Jr. "The Coalition for Networked Information's Acquisition-on-Demand Model: An Exploration and Critique." Serials Review 18, nos. 1-2 (1992): 78-81.

Bailey, Charles W., Jr. "Electronic (Online) Publishing in Action . . . The Public-Access Computer Systems Review and Other Electronic Serials." ONLINE 15 (January 1991): 28-35.

Bailey, Charles W., Jr. "Intelligent Multimedia Computer Systems: Emerging Information Resources in the Network Environment." Library Hi Tech 8, no. 1 (1990): 29-41.

Bailey, Charles W., Jr. "Network-Based Electronic Serials." Information Technology and Libraries 11 (March 1992): 29-35.

Bailey, Charles W., Jr., and Dana Rooks, eds. "Symposium on the Role of Network-Based Electronic Resources in Scholarly Communication and Research." The Public-Access Computer Systems Review 2, no. 2 (1991): 4-60. (To retrieve this article, send an e-mail message that says "GET BAILEY1 PRV2N2 F=MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)

Basch, Reva. "Books Online: Visions, Plans, and Perspectives for Electronic Text." ONLINE 15 (July 1991): 13-23.

Beazley, William G. "Impact of CALS on Electronic Publishing Systems and Users." Library Trends 38 (Spring 1990): 799-815.

Brown, Heather. "Standards for Structured Documents." The Computer Journal 32, no. 6 (1989): 505-514.

Bulick, Stephen. "Future Prospects for Network-Based Multimedia Information Retrieval." The Electronic Library 8 (April 1990): 88-99.

Butler, Brett. "Electronic Editions of Serials: The Virtual Library Model." Serials Review 18, nos. 1-2 (1992): 102-106.

Butler, Brett. "The Electronic Library Program: Developing Networked Electronic Library Collections." Library Hi Tech 9, no. 2 (1991): 21-30.

Cisler, Steve. "Convergent Electronic Cultures." Serials Review 18, nos. 1-2 (1992): 55-57.

Cline, Nancy. "Information Resources and the National Network." EDUCOM Review 25 (Summer 1990): 30-34.

+ Page 15 +

Dempsey, Lorcan. Libraries, Networks and OSI: A Review, with a Report on North American Developments. Bath, England: U.K. Office for Library Networking, the Library, University of Bath, 1991.

Dertouzos, Michael L. "Building the Information Marketplace." Technology Review 94 (January 1991): 29-40.

Drake, Miriam. "Buying Articles in the Future." Serials Review 18, nos. 1-2 (1992): 75-77.

Duggan, Mary Kay. "Copyright of Electronic Information: Issues and Questions." ONLINE 15 (May 1991): 20-26.

Garcia, Linda D. "Information Exchange: The Impact of Scholarly Communication." EDUCOM Review 25 (Fall 1990): 28-32.

Gardner, William. "The Electronic Archive: Scientific Publishing for the 1990s." Psychological Science 1 (November 1990): 333-341.

Getz, Malcolm. "Electronic Publishing: An Economic View." Serials Review 18, nos. 1-2 (1992): 25-31.

Grycz, Czeslaw Jan. "Economic Models for Networked Information." Serials Review 18, nos. 1-2 (1992): 11-18.

Hall, Stephen C. "The Four Stages of National Research and Education Network Growth." EDUCOM Review 26 (Spring 1991): 18-25.

Harnad, Stevan. "Interactive Publication: Extending the American Physical Society's Discipline-Specific Model for Electronic Publishing." Serials Review 18, nos. 1-2 (1992): 58-61.

Harnad, Stevan. "Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge." The Public-Access Computer Systems Review 2, no. 1 (1991): 39-53. (To retrieve this article, send an e-mail message that says "GET HARNAD PRV2N1 F=MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)

Harnad, Stevan. "Scholarly Skywriting and the Prepublication Continuum of Scientific Inquiry." Psychological Science 1 (November 1990): 342-344.

+ Page 16 +

Harrison, Teresa M., Timothy Stephen, and James Winter. "Online Journals: Disciplinary Designs for Electronic Scholarship." The Public-Access Computer Systems Review 2, no. 1 (1991): 25-38. (To retrieve this article, send an e-mail message that says "GET HARRISON PRV2N1 F=MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)

Hart, Michael S. "Project Gutenberg: Access to Electronic Texts." Database 13 (December 1990): 6-9.

Hickey, Thomas B., and Terry Noreault. "The Development of a Graphical User Interface for The Online Journal of Current Clinical Trials." The Public-Access Computer Systems Review 3, no. 2 (1992): 4-12. (To retrieve this article, send an e-mail message that says "GET HICKEY PRV3N2 F=MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)

Hugo, Jane, and Linda Newell. "New Horizons in Adult Education: The First Five Years (1987-1991)." The Public-Access Computer Systems Review 2, no. 1 (1991): 77-90. (To retrieve this article, send an e-mail message that says "GET HUGO PRV2N1 F=MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)

Jennings, Edward M. "EJournal: An Account of the First Two Years." The Public-Access Computer Systems Review 2, no. 1 (1991): 91-110. (To retrieve this article, send an e-mail message that says "GET JENNINGS PRV2N1 F=MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)

Jensen, Mary Brandt. "Making Copyright Work in Electronic Publishing Models." Serials Review 18, nos. 1-2 (1992): 62-65.

Kahle, Brewster, and Art Medlar. "An Information System for Corporate Users: Wide Area Information Servers." ONLINE 15 (September 1991): 56-60.

Kahn, Robert. "National Information Infrastructure Components." Serials Review 18, nos. 1-2 (1992): 85-87.

Kalin, Sally W., and Roy Tennant. "Beyond OPACS . . . The Wealth of Information Resources on the Internet." Database 14 (August 1991): 28-33.

Katz, Richard N. "Academic Information Management at the Crossroads: Time Again to Review the Economics." Serials Review 18, nos. 1-2 (1992): 41-44.

Kibbey, Mark, and Nancy H. Evans. "The Network is the Library." EDUCOM Review 24 (Fall 1989): 15-20.

+ Page 17 +

King, Tim. "Critical Issues for Providers of Network-Accessible Information." EDUCOM Review 26 (Summer 1991): 29-33.

King, Timothy B. "The Impact of Electronic and Networking Technologies on the Delivery of Scholarly Information." The Serials Librarian 21, nos. 2/3 (1991): 5-13.

Kost, Robert. "Technology Giveth . . ." Serials Review 18, nos. 1-2 (1992): 67-70.

Kovacs, Diane, Willard McCarty, and Michael Kovacs. "How to Start and Manage a BITNET LISTSERV Discussion Group: A Beginner's Guide." The Public-Access Computer Systems Review 2, no. 1 (1991): 128-143. (To retrieve this article, send an e-mail message that says "GET KOVACS PRV2N1 F=MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)

LaQuey, Tracy L., ed. The User's Directory of Computer Networks. Bedford, MA: Digital Press, 1990.

Laws, Kenneth I. "net.journalism." Serials Review 18, nos. 1-2 (1992): 82-84.

Lesk, Michael. "Pricing Electronic Information." Serials Review 18, nos. 1-2 (1992): 38-40.

Litchfield, Charles. "Local Storage and Retrieval of Electronic Journals: Training Issues for Technical Services Personnel." Serials Review 17, no. 4 (1991): 83-84.

Lucier, Richard E. "Knowledge Management: Refining Roles in Scientific Communication." EDUCOM Review 25 (Fall 1990): 21-27.

Lyman, Peter. "The Library of the (Not-so-Distant) Future." Change 23 (January/February 1991): 34-41.

Lynch, Clifford A. "Reaction, Response, and Realization: From the Crisis in Scholarly Communication to the Age of Networked Information." Serials Review 18, nos. 1-2 (1992): 107-112.

Lynch, Clifford A. "Visions of Electronic Libraries." In The Bowker Annual Library and Book Trade Almanac, 36th ed., comp. Filomena Simora, 75-82. New Providence, NJ: R. R. Bowker, 1991.

Lynch, Clifford A., and Cecilia M. Preston. "Internet Access to Information Resources." Annual Review of Information Science and Technology 25 (1990): 263-312.

+ Page 18 +

Maddox, John. "Electronic Journals Have a Future." Nature 356 (16 April 1992): 559.

Manoff, Marlene, Eileen Dorschner, Marilyn Geller, Keith Morgan, and Carter Snowden. "Report of the Electronic Journals Task Force MIT Libraries." Serials Review 18, nos. 1-2 (1992): 113-129.

McClure, Charles R., Ann P. Bishop, Philip Doty, and Howard Rosenbaum. The National Research and Education Network (NREN): Research and Policy Perspectives. Norwood, NJ: Ablex Publishing Corporation, 1991.

McMillan, Gail. "Embracing the Electronic Journal: One Library's Plan." The Serials Librarian 21, nos. 2/3 (1991): 97-108.

McMillan, Gail. "Technical Services for Electronic Journals Today." Serials Review 17, no. 4 (1991): 84-86.

Metz, Paul. "Electronic Journals from a Collection Manager's Point of View." Serials Review 17, no. 4 (1991): 82-83.

Metz, Paul, and Paul M. Gherman. "Serials Pricing and the Role of the Electronic Journal." College & Research Libraries 52 (July 1991): 315-327.

Okerson, Ann. "Back to Academia? The Case for American Universities to Publish Their Own Research." Logos 2, no. 2 (1991): 106-112.

Okerson, Ann. "The Electronic Journal: What, Whence, and When?" The Public-Access Computer Systems Review 2, no. 1 (1991): 5-24. (To retrieve this article, send an e-mail message that says "GET OKERSON PRV2N1 F=MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)

Okerson, Ann. "The Missing Model: A 'Circle of Gifts.'" Serials Review 18, nos. 1-2 (1992): 92-96.

Okerson, Ann. "Publishing Through the Network: The 1990s Debutante." Scholarly Publishing 23 (April 1992): 170-177.

Okerson, Ann. "With Feathers: Effects of Copyright and Ownership on Scholarly Publishing." College & Research Libraries 52 (September 1991): 425-438.

+ Page 19 +

Peters, Paul Evan. "Making the Market for Networked Information: An Introduction to a Proposed Program for Licensing Electronic Uses." Serials Review 18, nos. 1-2 (1992): 19-24.

Piternick, Anne B. "Electronic Serials: Realistic or Unrealistic Solution to the Journal 'Crisis'?" The Serials Librarian 21, nos. 2/3 (1991): 15-31.

Price-Wilkin, John. "Text Files in Libraries: Present Foundations and Future Directions." Library Hi Tech 9, no. 3 (1991): 7-44.

Quarterman, John S. The Matrix: Computer Networks and Conferencing Systems Worldwide. Bedford, MA: Digital Press, 1990.

Reich, Vicky. "Discipline-Specific Literature Bases: A View of the APS Model." Serials Review 18, nos. 1-2 (1992): 52-54, 65.

Rogers, Sharon J., and Charlene S. Hurt. "How Scholarly Communication Should Work in the 21st Century." College & Research Libraries 51 (January 1990): 5-8.

Savage, Lon. "The Journal of the International Academy of Hospitality Research." The Public-Access Computer Systems Review 2, no. 1 (1991): 54-66. (To retrieve this article, send an e-mail message that says "GET SAVAGE PRV2N1 F=MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)

Schultz, T. D. "A World Physics Information System: An Online, Highly Interactive, Discipline-Oriented Facility." Serials Review 18, nos. 1-2 (1992): 45-48.

Strangelove, Michael, and Diane Kovacs. Directory of Electronic Journals, Newsletters and Academic Discussion Lists, 2nd ed. Washington, D.C.: Office of Scientific and Academic Publishing, Association of Research Libraries, 1992.

"Task Force Report Looks at Future of Information Services." Bulletin of the American Physical Society 36 (April 1991): 1105-1151.

+ Page 20 +

Tuttle, Marcia. "The Newsletter on Serials Pricing Issues." The Public-Access Computer Systems Review 2, no. 1 (1991): 111-127.
(To retrieve this article, send an e-mail message that says "GET TUTTLE PRV2N1 F=MAIL" to LISTSERV@UHUPVM1 or LISTSERV@UHUPVM1.UH.EDU.)

Yankelovich, Nicole. "Three Pieces of the Puzzle: Wide-Area Hypermedia, Information Agents and On-line Reference Works." Psychological Science 1 (November 1990): 350-352.

Yavarkovsky, Jerome. "A University-Based Electronic Publishing Network." EDUCOM Review 25 (Fall 1990): 14-20.

Young, Peter R. "National Corporation for Scholarly Publishing: Presentation and Description of the Model." Serials Review 18, nos. 1-2 (1992): 100-101.

The Public-Access Computer Systems Review is an electronic journal that is distributed on BITNET, Internet, and other computer networks. There is no subscription fee.

To subscribe, send an e-mail message to LISTSERV@UHUPVM1 (BITNET) or LISTSERV@UHUPVM1.UH.EDU (Internet) that says: SUBSCRIBE PACS-P First Name Last Name. PACS-P subscribers also receive two electronic newsletters: Current Cites and Public-Access Computer Systems News.

This article is Copyright (C) 1992 by Charles W. Bailey, Jr. All Rights Reserved.

The Public-Access Computer Systems Review is Copyright (C) 1992 by the University Libraries, University of Houston. All Rights Reserved.

Copying is permitted for noncommercial use by computer conferences, individual scholars, and libraries. Libraries are authorized to add the journal to their collection, in electronic or printed form, at no charge. This message must appear on all copied material. All commercial use requires permission.
